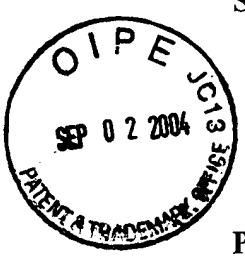
AF FFW



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Inventor(s): G. J. Foschini
A. Lozano
F. Rashid-Farrokhi
R. A. Valenzuela

Case: 10-1-3-14
Serial No.: 09/482,429 **Group Art Unit:** 2666
Filed: January 13, 2000
Examiner: F. Duong
Title: Space-Time Processing For Multiple-Input, Multiple-Output, Wireless Systems

THE COMMISSIONER OF PATENTS AND TRADEMARKS
WASHINGTON, D. C. 20231

SIR:

Appellant's Brief Under 37 C.F.R. 1.192

This is an appeal to the Board of Patent Appeals and Interferences from the Final Rejection dated April 5, 2004. Applicants are submitting this Brief in triplicate.

A Notice of Appeal was timely filed.

Real Party in Interest

The real party in interest is Lucent Technologies Inc.

Related Appeals and Interferences

There are no related appeals or interferences.

Status of Claims

Claims 1-22 are pending in the application.

Claims 21 and 22 stand allowed.

Claims 5, 15, and 16 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Claims 1-4, 6-14, and 17-20 are rejected based on prior art.

A copy of the claims under appeal as now presented are appended to this brief in Appendix A.

Status of Amendments

All amendments to the claims have been entered.

Summary of the Invention

It is well known in the art that multiple-input, multiple-output (MIMO) systems can achieve dramatically improved capacity as compared to single antenna, i.e., single antenna to single antenna or multiple antenna to single antenna, systems. However, to achieve this improvement, it is preferable that there be a rich scattering environments, so that the various signals reaching the multiple receive antennas be largely uncorrelated. If the signals have some degree of correlation, and such correlation is ignored, performance degrades and capacity is reduced.

The invention relates to a way of developing signals in a MIMO system such that, even in the face of some correlation, the most performance and capacity that can be achieved with a channel of that level of correlation is obtained. In accordance with the principles of the invention, the signals transmitted from the various antennas are processed so as to improve the ability of the receiver to extract them from the received signal. More specifically the number of bit streams that is transmitted simultaneously is adjusted, e.g., reduced, depending on the level of correlation, while multiple versions of each bit stream, variously weighted, are transmitted simultaneously. The variously weighted versions are combined to produced one combined weighted signal, a so-called "transmit vector", for each antenna. The receiver processes the received signals in the same manner as it would have had all the signals reaching the receive antennas been uncorrelated.

In one embodiment of the invention, the weight vectors are determined by the forward channel transmitter using the channel properties of the forward link which are

made known to the transmitter of the forward link by being transmitted from the receiver of the forward link by the transmitter of the reverse link. In another embodiment of the invention the weight vectors are determined by the forward channel receiver using the channel properties of the forward link and the determined weight vectors are made known to the transmitter of the forward link by being transmitted from the receiver of the forward link by the transmitter of the reverse link.

The channel properties used to determine the weight vectors includes the channel response from the transmitter to the receiver and the covariance matrix of noise and interference, e.g., the interference covariance matrix, measured at the receiver.

Issues

I. Are claims 17-19 properly rejected under 35 U.S.C. 102(b) as being anticipated by EP 0 807 989 A1.

II. Are claims 1-4 and 6-14 properly rejected under 35 U.S.C. 103(a) as being unpatentable over United States Patent No. 6,351,499 issued to Paulraj et al. on February 26, 2002 in view of United States Patent No. 5,982,327 issued to Vook et al. on November 9, 1999.

Grouping of Claims

Claims 1-5 are a group of method claims.

Claims 6-9 are a group of apparatus claims.

Claims 10-16 are a group of apparatus claims.

Claim 17 is an apparatus claim that is a group unto itself.

Claim 18 is an apparatus claim that is a group unto itself.

Claims 19 and 20 are a group of apparatus clams.

Claims 21 and 22 are each a separate group that stands allowed.

Claims 5, 15, and 16 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Each group of claims, as well as each objected to claim, stand separately.

Arguments

Issues I and II – Prior-Art-Based Rejections

Claims 17-19 are rejected under 35 U.S.C. 102(b) as being anticipated by EP 0 807 989 A1.

Claims 1-4 and 6-14 are rejected under 35 U.S.C. 103(a) as being unpatentable over United States Patent No. 6,351,499 issued to Paulraj et al. on February 26, 2002 in view of United States Patent No. 5,982,327 issued to Vook et al. on November 9, 1999.

These grounds of rejections are respectfully traversed for the following reasons.

Each claim that is rejected based on prior art recites the use of an interference covariance matrix in the use of the computation of weights. The cited references do **not** teach employing an interference covariance matrix in computing weights.

Applicants note that some of the references compute the covariance matrix of signals. However, notwithstanding the Office Action's suggestion to the contrary, the covariance matrix of signals is quite different from the interference covariance matrix. Use of the interference covariance matrix, as recited in applicants' claims, advantageously, allows the invention to be employed in an environment in which the noise plus interference is not white, the interference contributing the not white component. This is not possible with the cited references.

Nevertheless, the Final and Advisory Office Actions maintains that in teaching to use the covariance matrix of signals in computing weights the prior art already teaches applicants' recited limitation of employing an interference covariance matrix in computing weights. This is based, according to the Office Action, on giving the limiting term "interference covariance matrix" what the Office Action calls its broadest reasonable interpretation, since the term is not defined in the specification. Apparently then, according to the Office Action's broad interpretation, the term "interference covariance matrix" is so broad that it includes the "covariance matrix of signals".

The Office Action's analysis and conclusion are, quite simply, incorrect.

The Office Action states that it recognizes that applicants may be their own lexicographer. However, applicants are not defining any terms, and instead are using the

terms in a well-accepted manner. Instead, it is the Office Action that acting as a lexicographer, in a manner that is **not** permitted. Indeed, the Office Action is attempting to state, analogously, that when applicants recite "red", they really mean "green".

As is well known, in communication systems, "signals" are the desired information to be communicated, while "interference" and "noise", are the undesirable effects imparted by a channel onto the signal. Interference is caused by the signals of others, while noise is caused by nature, e.g., the motion of electrons as a result of temperature. As a result, noise is typically "white" in space, time and frequency, while interference generally is not "white".

Thus, applicants' interference covariance matrix relates to the undesired. By contrast, the Office Action's covariance matrix of signals relates to the desired. It is unreasonable to say that they are the same thing, or even that one suggests the other.

Applicant's note that the term "interference covariance matrix", a.k.a., "covariance matrix of interference", and the term "covariance matrix of signals" are terms of art. More specifically, the interference covariance matrix results from applying basic mathematics to interference encountered in a MIMO communications system. As such, "covariance matrix of signals", as used in the references, can**not** be defined to include the limitation of "interference covariance matrix" as recited in applicants' claims. This can be clearly seen from the following basic explanation.

Mathematically, typically, a random vector has a covariance matrix associated with it. The covariance matrix for a vector is a generalization of the concept of a variance for a random scalar. More specifically, the i,j entry of the covariance matrix indicates the cross-variance between the i^{th} and the j^{th} entries of the vector to which the covariance pertains.

In MIMO communication systems, the signal, i.e., the desired information, and the interference, i.e., the other undesired stuff, i.e., besides the noise, that gets introduced by the channel and causes the signal to be degraded, are, naturally, separate items and each is represented by its own distinct vector. In a MIMO system, the signal is represented as a vector and the interference is represented by its own separate vector. That the signal is represented as a vector and the interference is also represented by a

vector is due to the multidimensional nature of the MIMO system. Naturally then, the signal vector and the interference vector each has associated with it its own respective covariance matrix.

In other words, given that there is a signal vector, there is a covariance matrix for that signal vector, known as the signal covariance matrix. Also, given that there is an interference vector, there is a covariance matrix for the interference vector, known as the interference covariance matrix.

Thus, notwithstanding the Office Action's suggestion to the contrary, the signal covariance matrix and the interference covariance matrix are different quantities and they are respectively computed based on different vectors.

Clearly then, the two terms cannot be simply substituted for one another. One relates to the signal, which is good, the other to the interference, which is bad. Furthermore, use of one does not suggest use of the other. In fact, the cited references do not treat the interference as a vector, but instead, simplify its representation to be a scalar. Thus, the references, at best, have covariance matrix of signals, but they do not have an interference covariance matrix.

Attention is directed to the paper, attached hereto, entitled *MIMO Channel Capacity in Co-Channel Interference* by Yi Song and Steven D. Blostein, in which equation 2 shows the covariance matrix of the interference-plus-noise and equation 3 shows the covariance matrix of the received signal, thus clearly distinguishing between the two. Note that the recited term interference covariance matrix is the covariance matrix of the interference-plus-noise when the noise is negligible with respect to the interference.

Other articles, attached hereto as well, which use the term interference covariance matrix as a term of art are, for example, *Channel Estimation and Data Detection for MIMO Systems under Spatially and Temporally Colored Interference* by Yi Song and Steven D. Blostein; *Unique Features of Subspace Processing for Adaptive Radar in Inhomogeneous Environments* by William L. Melvin; *Interference Mitigation in STAP Using the Two-Dimensional Wold Decomposition Model* by Joseph M. Francos; and

Subspace Approximation for Adaptive Multichannel Radar Filtering by A.W. Bojanczyk, W.L. Melvin, and E.J. Holder

Furthermore, the use of an interference covariance matrix is **not** obvious from the use of a scalar to represent interference. Indeed, even if it were suggested, which it is not, making use of the interference covariance matrix is not simple calculation, nor is it a straightforward substitution of a matrix for a scalar.

Thus, applicants use of the interference covariance matrix, and recitation of same in the claims, renders applicants' invention as claimed patentable over the cited references.

The Office Action appears to be looking for some sort of explicit claim language that states that the interference-plus-noise being dealt with by applicants is not white. Not having found such language, the Office Action states that there is no such distinguishing limitation in the claims. However, applicants note that it is the use of the interference covariance matrix that allows the invention to be employed in an environment in which the interference is not white, and use of the interference covariance matrix is indeed recited in applicants' claims. Applicants in their previous response were simply pointing out a distinguishing advantage that results from the invention as claimed.

Thus, all of applicants' claims are allowable over the cited references, individually or in combination.

Conclusion

In view of the foregoing, it is submitted that the Examiner is in error. It is, accordingly, respectfully requested that the rejection of claims 1-4, 6-14, and 17-20 be reversed and the application passed to issue.

Respectfully,

G. J. Foschini
A. Lozano
F. Rashid-Farrokhi
R. A. Valenzuela

By Eugene J. Rosenthal

Eugene J. Rosenthal, Attorney
Reg. No. 36,658
732-949-1857

Lucent Technologies Inc.

Date: 8/31/04

APPENDIX A

Claims

1. A method for transmitting signals in communications system having a transmitter with N transmit antennas transmitting over a forward channel to a receiver having L receiver antennas and a reverse channel for communicating from said receiver to said transmitter, in which there may exist correlation in the signals received by two or more of said L receive antennas, the method comprising the steps of:

determining the number of independent signals that can be transmitted from said N transmit antennas to said L receive antennas;

creating, from a data stream, a data substream to be transmitted for each of the number of independent signals that can be transmitted from said N transmit antennas to said L receive antennas;

weighting each of said substreams with N weights, one weight for each of said N transmit antennas, said weights being determined by said transmitter as a function of channel information and an interference covariance matrix, to produce N weighted substreams per substream;

combining one of said weighted substreams produced from each of said substreams for each of said transmit antennas to produce a transmit signal for each of said transmit antennas.

2. The invention as defined in claim 1 further comprising the step of transmitting said transmit signal from a respective one of said antennas.

3. The invention as defined in claim 1 further comprising the step of receiving said weights via said reverse channel.

4. The invention as defined in claim 1 wherein said channel information and said interference covariance matrix are received by said transmitter from said receiver via said reverse channel.

5. The invention as defined in claim 1 wherein said weights are determined by solving a matrix equation $H^{\dagger}(K^{N})H = U^{\dagger}\Lambda^{2}U$ where:

H is a channel response matrix,

$H^{\dagger}$ is a conjugate transpose of said channel response matrix H,

K^{N} is the interference covariance matrix,

U is a unitary matrix, each column of which is an eigenvector of $H^{\dagger}(K^{N})H$,

Λ is a diagonal matrix defined as $\Lambda = diag(\lambda^{1},\ldots,\lambda^{M})$, where $\lambda^{1},\ldots,\lambda^{M}$ are each eignevalues of $H^{\dagger}(K^{N})H$, M being the maximum number of nonzero eigenvalues, which corresponds to the number of said independent signals, and

$U^{\dagger}$ is the conjugate transpose of matrix U;

waterfilling said eigenvalues λ by solving the simultaneous equations $\tilde{\lambda}^{k} = (\nu - \frac{1}{(\lambda^{k})^{2}})^{+}$ and $\sum_{k} \tilde{\lambda}^{k} = P$, for ν, where:

k is an integer index that ranges from 1 to M,

P is the transmitted power,

+ is an operator that returns zero (0) when its argument is negative, and returns the argument itself when it is positive, and

each $\tilde{\lambda}$ is an intermediate variable representative of a power for each weight vector;

defining matrix Φ as $\Phi = U^{\dagger} diag(\tilde{\lambda}^{1},\ldots,\tilde{\lambda}^{M})U$, where *diag* indicates that the various $\tilde{\lambda}$ are arranged as the elements of the main diagonal of matrix Φ;

wherein each column of matrix Φ is used as a normalized weight vector indicated by $\Phi = [z_{1},\ldots,z_{N}]$ and said normalized weight vectors are made up of individual normalized weights z, $z_{i} = [z_{i1},\ldots,z_{iN}]$, where i is an integer ranging from 1 to N;

developing an unnormalized weight vector $w_{i} = [w_{i1},\ldots,w_{iN}]$, with each of said weights therein being $\sqrt{\tilde{\lambda}^{i}}\, z_{ij}$, where j is an integer ranging from1 to N.

6. Apparatus for transmitting signals in communications system having a transmitter with N transmit antennas transmitting over a forward channel to a receiver having L receiver antennas and a reverse channel for communicating from said receiver to said transmitter, in which there may exist correlation in the signals received by two or more of said L receive antennas, the apparatus comprising:

means for determining the number of independent signals that can be transmitted from said N transmit antennas to said L receive antennas;

means for creating, from a data stream, a data substream to be transmitted for each of the number of independent signals that can be transmitted from said N transmit antennas to said L receive antennas;

means for weighting each of said substreams with N weights, one weight for each of said N transmit antennas, said weights being determined by said apparatus for transmitting signals as a function of information about said forward channel and an interference covariance matrix, to produce N weighted substreams per substream;

means for combining one of said weighted substreams produced from each of said substreams for each of said antennas to produce a transmit signal for each antenna.

7. The invention as defined in claim 6 wherein said transmitter comprises means for developing said weights.

8. The invention as defined in claim 6 wherein said transmitter comprises means for storing said weights.

9. The invention as defined in claim 6 wherein said receiver comprises means for developing said weights.

10. A transmitter for transmitting signals in communications system having a transmitter with N transmit antennas transmitting over a forward channel to a receiver having L receiver antennas and a reverse channel for communicating from said receiver to said transmitter, in which there may exist correlation in the signals received by two or more of said L receive antennas, the transmitter comprising:

a demultiplexor for creating, from a data stream, a data substream to be transmitted for each of the number of independent signals that can be transmitted from said N transmit antennas to said L receive antennas

multipliers for weighting each of said substreams with N weights, one weight for each of said N transmit antennas, wherein said weights are determined in said transmitter in response to an interference covariance matrix estimate and an estimate of the forward channel response, to produce N weighted substreams per substream, each of said weights being a function of at least an estimate interference covariance matrix and an estimate of a forward matrix channel response between said transmitter and said receiver; and

adders for combining one of said weighted substreams produced from each of said substreams for each of said antennas to produce a transmit signal for each of said transmit antennas.

11. The invention as defined in claim 10 further comprising a digital to analog converter for converting each of said combined weighted substreams.

12. The invention as defined in claim 10 further comprising an upconverter for converting to radio frequencies each of said analog-converted combined weighted substreams.

13. The invention as defined in claim 10 wherein said interference covariance matrix estimate and said estimate of the forward channel response are received by said transmitter from said receiver over said reverse channel.

14. The invention as defined in claim 10 wherein said weights are determined in said receiver and are transmitted to said transmitter over said reverse channel.

15. The invention as defined in claim 10 wherein said weights are determined by solving a matrix equation $H^{\dagger}(K^{N})H = U^{\dagger}\Lambda^{2}U$ where:

H is a channel response matrix,

$H^{\dagger}$ is a conjugate transpose of said channel response matrix H,

K^N is the interference covariance matrix,

U is a unitary matrix, each column of which is an eigenvector of $H^{\dagger}(K^N)H$,

Λ is a diagonal matrix defined as $\Lambda = diag(\lambda^1,\ldots,\lambda^M)$, where $\lambda^1,\ldots,\lambda^M$ are each eignevalues of $H^{\dagger}(K^N)H$, M being the maximum number of nonzero eigenvalues, which corresponds to the number of said independent signals, and

$U^{\dagger}$ is the conjugate transpose of matrix U;

waterfilling said eigenvalues λ by solving the simultaneous equations $\tilde{\lambda}^k = (\nu - \frac{1}{(\lambda^k)^2})^+$ and $\sum_k \tilde{\lambda}^k = P$, for ν, where:

k is an integer index that ranges from 1 to M,

P is the transmitted power,

+ is an operator that returns zero (0) when its argument is negative, and returns the argument itself when it is positive, and

each $\tilde{\lambda}$ is an intermediate variable representative of a power for each weight vector;

defining matrix Φ as $\Phi = U^{\dagger} diag(\tilde{\lambda}^1,\ldots,\tilde{\lambda}^M)U$, where *diag* indicates that the various $\tilde{\lambda}$ are arranged as the elements of the main diagonal of matrix Φ;

wherein each column of matrix Φ is used as a normalized weight vector indicated by $\Phi = [z_1,\ldots,z_N]$ and said normalized weight vectors are made up of individual normalized weights z, $z_i = [z_{i1},\ldots,z_{iN}]$, where i is an integer ranging from 1 to N;

developing unnormalized weight vector $w_i = [w_{i1},\ldots,w_{iN}]$, with each of said weights therein being $\sqrt{\tilde{\lambda}^i} z_{ij}$, where j is an integer ranging from1 to N.

16. The invention as defined in claim 10 wherein said transmitter and receiver communicate using time division multiplexing (TDD) and said weights are determined in said transmitter using an estimate of the forward channel response that is determined by a receiver of said reverse link for said transmitter.

17. A receiver for use in a MIMO system, comprising:

L antennas;

L downconverters;

an estimator for determining an estimate of an interference covariance matrix for a forward channel being received by said receiver; and

a transmitter for a reverse channel for transmitting said estimate of an interference covariance matrix to a receiver for said reverse channel.

18. A receiver for use in a MIMO system, comprising:

L antennas;

L downconverters;

an estimator for determining an estimate of an interference covariance matrix for a forward channel being received by said receiver;

an estimator for determining an estimate of a channel response for a forward channel being received by said receiver; and

a transmitter for a reverse channel for transmitting said estimate of an interference covariance matrix and said estimate of a channel response to a receiver for said reverse channel.

19. A receiver for use in a MIMO system, comprising:

an estimator for determining an estimate of an interference covariance matrix for a forward channel being received by said receiver;

an estimator for determining an estimate of a channel response for a forward channel being received by said receiver; and

a weight calculator for calculating weights for use by a transmitter of said forward channel to transmit data substreams to said receiver as a function of said estimate of an interference covariance matrix for a forward channel being received by said receiver and said estimate of a channel response for a forward channel being received by said receiver.

20. The invention as defined in claim 19 further including a transmitter for a reverse channel for transmitting said weights to a receiver for said reverse channel.

21. A receiver for use in a MIMO system, comprising:

L antennas;

L downconverters;

an estimator for determining an estimate of an interference covariance matrix for a forward channel being received by said receiver;

an estimator for determining an estimate of a channel response for a forward channel being received by said receiver; and

a weight calculator for calculating weights for use by a transmitter of said forward channel to transmit data substreams to said receiver, said weights being determined in said weight calculator by

solving a matrix equation $H^{\dagger}(K^{N})H = U^{\dagger}\Lambda^{2}U$ where:

H is a channel response matrix,

$H^{\dagger}$ is a conjugate transpose of said channel response matrix H,

K^N is the interference covariance matrix,

U is a unitary matrix, each column of which is an eigenvector of $H^{\dagger}(K^N)H$,

Λ is a diagonal matrix defined as $\Lambda = diag(\lambda^1,\ldots,\lambda^M)$, where $\lambda^1,\ldots,\lambda^M$ are each eignevalues of $H^{\dagger}(K^N)H$, M being the maximum number of nonzero eigenvalues, which corresponds to the number of said independent signals, and

$U^{\dagger}$ is the conjugate transpose of matrix U;

waterfilling said eigenvalues λ by solving the simultaneous equations $\tilde{\lambda}^k = (\nu - \frac{1}{(\lambda^k)^2})^+$ and $\sum_k \tilde{\lambda}^k = P$, for ν, where:

k is an integer index that ranges from 1 to M,

P is the transmitted power,

+ is an operator that returns zero (0) when its argument is negative, and returns the argument itself when it is positive, and

each $\tilde{\lambda}$ is an intermediate variable representative of a power for each weight vector;

defining matrix Φ as $\Phi = U^{\dagger} diag(\tilde{\lambda}^1,\ldots,\tilde{\lambda}^M)U$, where *diag* indicates that the various $\tilde{\lambda}$ are arranged as the elements of the main diagonal of matrix Φ;

wherein each column of matrix Φ is used as a normalized weight vector indicated by $\Phi = [z_1,\ldots,z_N]$ and said normalized weight vectors are made up of individual normalized weights z, $z_i = [z_{i1},\ldots,z_{iN}]$, where i is an integer ranging from 1 to N;

developing unnormalized weight vector $w_i = [w_{i1},\ldots,w_{iN}]$, with each of said weights therein being $\sqrt{\tilde{\lambda}^i} z_{ij}$, where j is an integer ranging from1 to N.

22. A method for determining weights for use in transmitting signals in communications system having a transmitter with N transmit antennas transmitting over a forward channel to a receiver having L receiver antennas and a reverse channel for communicating from said receiver to said transmitter, in which there may exist correlation in the signals received by two or more of said L receive antennas, the method comprising the steps of:

determining the number of independent signals M that can be transmitted from said N transmit antennas to said L receive antennas through a process of determining weights for substreams derived from data to be transmitted via said N antennas as part of forming said signals, wherein said weights are determined by

solving a matrix equation $H^{\dagger}(K^N)H = U^{\dagger}\Lambda^2 U$ where:

H is a channel response matrix,

$H^{\dagger}$ is a conjugate transpose of said channel response matrix H,

K^N is the interference covariance matrix,

U is a unitary matrix, each column of which is an eigenvector of $H^{\dagger}(K^N)H$,

Λ is a diagonal matrix defined as $\Lambda = diag(\lambda^1,\ldots,\lambda^M)$, where $\lambda^1,\ldots,\lambda^M$ are each eigenvalues of $H^{\dagger}(K^N)H$, M being the maximum number of nonzero eigenvalues, which corresponds to the number of said independent signals, and

$U^{\dagger}$ is the conjugate transpose of matrix U;

waterfilling said eigenvalues λ by solving the simultaneous equations $\tilde{\lambda}^k = (\nu - \frac{1}{(\lambda^k)^2})^+$ and $\sum_k \tilde{\lambda}^k = P$, for ν, where:

k is an integer index that ranges from 1 to M,

P is the transmitted power,

+ is an operator that returns zero (0) when its argument is negative, and returns the argument itself when it is positive, and

each $\tilde{\lambda}$ is an intermediate variable representative of a power for each weight vector;

defining matrix Φ as $\Phi = U^{\dagger} diag(\tilde{\lambda}^1,\ldots,\tilde{\lambda}^M)U$, where *diag* indicates that the various $\tilde{\lambda}$ are arranged as the elements of the main diagonal of matrix Φ;

wherein each column of matrix Φ is used as a normalized weight vector indicated by $\Phi = [z_1,\ldots,z_N]$ and said normalized weight vectors are made up of individual normalized weights z, $z_i = [z_{i1},\ldots,z_{iN}]$, where i is an integer ranging from 1 to N;

developing unnormalized weight vector $w_i = [w_{i1},\ldots,w_{iN}]$, with each of said weights therein being $\sqrt{\tilde{\lambda}^i} z_{ij}$, where j is an integer ranging from1 to N.

MIMO Channel Capacity in Co-Channel Interference

Yi Song and Steven D. Blostein
Department of Electrical and Computer Engineering
Queen's University
Kingston, Ontario, Canada, K7L 3N6
E-mail: {songy, sdb}@ee.queensu.ca

Abstract— **Recent information theory results have indicated that a large channel capacity exists for wireless systems with multiple transmit and receive antennas. With different assumptions of channel knowledge and interference knowledge at the transmitter, the channel capacity of multiple input multiple output (MIMO) systems has been studied under both spatially white and colored interference and noise. In this paper, we fix the total interference-plus-noise power and evaluate the outage capacity under two spatially colored interference environments: (1) a few high-data-rate interferers each with high power, (2) a large number of low-data-rate interferers each with low power. The results show that MIMO capacity is larger with fewer high-data-rate interferers. We also assess the impact of an estimated channel and/or interference on capacity. In the case of 4 transmit and 4 receive antennas for the user of interest, 10 interferers, total-interference-to-noise ratio and signal-to-noise ratio are both 20dB, the results show that it is beneficial to estimate the channel and/or interference if the variance of estimation error is less than about 50% of the variance of true channel and/or interference.**

I. Introduction

Recent information theory results have indicated that a large channel capacity exists for wireless systems with multiple transmit and receive antennas [1]. With different assumptions of channel knowledge and interference knowledge at the transmitter, the channel capacity of multiple input multiple output (MIMO) systems have been studied under both spatially white and colored interference and noise by applying different power allocation schemes at the transmitter [2][3]. Meanwhile, in future generation wireless communication systems, multi-rate data services will be dominant. To support users of different data rate at a certain quality of service (e.g., a certain level of bit error rate), the user's transmit power is, in general, proportional to the data rate. Therefore, high-data-rate users need high transmit powers.

In multiple-access systems, the interference is, in general, spatially colored. In this paper, we fix the total interference-plus-noise power and examine the MIMO outage capacity under two spatially colored interference environments: (1) a few high-data-rate interferers each with high power, (2) a large number of low-data-rate interferers each with low power. We would like to find out under which interference environment a MIMO system achieves a higher outage capacity. The assumption of fixed total interference-plus-noise power is reasonable since in wireless systems, likely there is some mechanism, such as power control, to control the interference experienced by a user. The result of this work has the implication on design of the medium access control (MAC) protocols and scheduling of packet transmissions in future wireless systems. We will also assess the impact of an estimated channel and/or interference on capacity.

II. System Model

We consider a single-user narrowband link with interference from other users. The user of interest is equipped with M transmit and N receive antennas. Each interfering user has one transmit antenna, and the same N receive antennas as the user of interest. The received signal vector $\mathbf{y}$ $(N \times 1)$ is

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \underbrace{\sqrt{\frac{P_I}{L}} \sum_{i=1}^{L} \mathbf{h}_i s_i + \mathbf{w}}_{\mathbf{n}} \tag{1}$$

where $\mathbf{H}$ $(N \times M)$ is the MIMO channel matrix of the user of interest, $\mathbf{s}$ $(M \times 1)$ is the transmit signal vector of the user of interest, $\mathbf{n}$ $(N \times 1)$ is the interference-plus-noise vector at the receiver. The number of interferers is L, P_I is the fixed total interference power, $\mathbf{h}_i$ $(N \times 1)$ is the channel vector of the ith interferer, s_i is the ith interferer's transmit signal with unit power, and $\mathbf{w}$ $(N \times 1)$ is the thermal noise with covariance matrix $\mathbf{E}\{\mathbf{w}\mathbf{w}^\dagger\} = \sigma^2\mathbf{I}_N$ where $\dagger$ denotes transpose conjugate. The channel matrix $\mathbf{H}$ and the channel vectors $\mathbf{h}_i$'s are mutually independent, and assumed to be quasi-static (constant over one frame) having uncorrelated realizations in different frames. It is further assumed that the elements in $\mathbf{H}$ and $\mathbf{h}_i$ are identically independent distributed (i.i.d.) complex Gaussian random variables (RVs) with zero mean and unit variance. This implies flat Rayleigh fading and that antennas are separated far apart. The signal of the user of interest $\mathbf{s}$, the interfering signal $\mathbf{s}_i$, and the thermal noise $\mathbf{w}$ are mutually independent. It is obvious in (1) that each interferer has the same power. More interferers in the system, lower power each interferer has. It can be shown that the covariance matrix of the interference-plus-noise is

$$\mathbf{R}_0 = \mathbf{E}\{\mathbf{n}\mathbf{n}^\dagger\} = \frac{P_I}{L} \sum_{i=1}^{L} \mathbf{h}_i\mathbf{h}_i^\dagger + \sigma^2\mathbf{I}_N, \tag{2}$$

and the covariance matrix of the received signal is

$$\mathbf{E}\{\mathbf{y}\mathbf{y}^\dagger\} = \mathbf{H}\Sigma_s\mathbf{H}^\dagger + \mathbf{R}_0 \tag{3}$$

where $\Sigma_s = \mathbf{E}\{\mathbf{s}\mathbf{s}^\dagger\}$.

In our system model, we assume each interferer has one transmit antenna. However, it is easy to accommodate interferers with more than one transmit antenna by aggregating several interfering users with one transmit antenna.

III. Channel Capacity

In this section, we derive the MIMO channel capacity with spatially colored interference and under different assumptions of channel and interference knowledge at the transmitter: (1) both channel and interference covariance matrices $\mathbf{H}$ and $\mathbf{R_0}$ are available, (2) only $\mathbf{H}$ is available, and (3) neither $\mathbf{H}$ nor $\mathbf{R_0}$ is available. In all the cases, we assume that the receiver knows the channel $\mathbf{H}$. Comparing to [3], our derivation uses a modeled interference covariance matrix as (2). In addition, we give a new interpretation of MIMO channel capacity under spatially colored interference.

We introduce the differential entropy of a circularly symmetric complex Gaussian random vector. If $\mathbf{x}$ is a circularly symmetric complex Gaussian random vector with covariance matrix $\mathbf{Q}$, the differential entropy of $\mathbf{x}$ is $\log_2 \det(\pi e \mathbf{Q})$. In addition, circularly symmetric complex Gaussians are entropy maximizers [4].

Assuming the interference-plus-noise $\mathbf{n}$ in (1) is circularly symmetric complex Gaussian, the optimal distribution for the signal $\mathbf{s}$ is then circularly symmetric complex Gaussian [4] [5]. As the receiver knows the channel, the mutual information between the channel input and output is given as

$$\begin{aligned} &\mathcal{I}(\mathbf{s};\mathbf{y}) \\ &= \log_2 \det\left[\pi e(\mathbf{H}\Sigma_s\mathbf{H}^\dagger + \mathbf{R_0})\right] - \log_2 \det(\pi e \mathbf{R_0}) \\ &= \log_2 \det\left(\mathbf{I}_N + \mathbf{R}^{-1}\mathbf{H}\frac{\Sigma_s}{\sigma^2}\mathbf{H}^\dagger\right) \\ &= \log_2 \det\left[\mathbf{I}_N + \left(\mathbf{R}^{-1/2}\mathbf{H}\right)\frac{\Sigma_s}{\sigma^2}\left(\mathbf{R}^{-1/2}\mathbf{H}\right)^\dagger\right] \end{aligned} \tag{4}$$

where

$$\mathbf{R} = \frac{P_I}{\sigma^2 \cdot L}\sum_{i=1}^{L}\mathbf{h}_i\mathbf{h}_i^\dagger + \mathbf{I}_N, \tag{5}$$

and the third equality comes from the facts that $\det(\mathbf{I} + AB) = \det(\mathbf{I}+BA)$ for square matrices A and B and $\mathbf{R}^{-1/2}$ is Hermitian. We denote $\frac{P_I}{\sigma^2}$ as the ratio of total interference power to noise power. The channel capacity is the maximized mutual information with transmit power constraint $tr(\Sigma_s) \le P_T$, i.e.,

$$C = \max_{tr(\frac{\Sigma_s}{\sigma^2}) \le \frac{P_T}{\sigma^2}} \log_2 \det\left[\mathbf{I}_N + \left(\mathbf{R}^{-1/2}\mathbf{H}\right)\frac{\Sigma_s}{\sigma^2}\left(\mathbf{R}^{-1/2}\mathbf{H}\right)^\dagger\right], \tag{6}$$

where $\frac{P_T}{\sigma^2}$ is the ratio of signal power to noise power.

Eqn. (6) suggests that we could consider $\mathbf{R}^{-1/2}\mathbf{H}$ as a combined channel. As a result, the capacity in (6) is equivalent to the capacity of the combined channel $\mathbf{R}^{-1/2}\mathbf{H}$ under spatially white noise. With this new interpretation and the results of channel capacity under spatially white noise in [2], we obtain the capacity with spatially colored interference and noise.

A. Both channel and interference information at the transmitter

By applying water-filling power allocation with the combined channel $\mathbf{R}^{-1/2}\mathbf{H}$ at the transmitter, the channel capacity is

$$C = \sum_{i=1}^{M}\log_2(1 + p_i\lambda_i), \tag{7}$$

the optimal transmit signal covariance matrix is

$$\Sigma_s = \sigma^2\mathbf{U}\,\mathrm{diag}(p_1, \ldots, p_M)\mathbf{U}^\dagger \tag{8}$$

where

$$\mathbf{H}^\dagger\mathbf{R}^{-1}\mathbf{H} = \mathbf{U}\Lambda\mathbf{U}^\dagger, \quad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_M), \tag{9}$$

$\lambda_1, \ldots, \lambda_M$ are the eigenvalues of $\mathbf{H}^\dagger\mathbf{R}^{-1}\mathbf{H}$, $\mathbf{U}$ is an unitary matrix consisting of eigenvectors of $\mathbf{H}^\dagger\mathbf{R}^{-1}\mathbf{H}$,

$$p_i = (\mu - \frac{1}{\lambda_i})^+ \tag{10}$$

where μ is chosen such that

$$\sum_{i=1}^{M} p_i = \frac{P_T}{\sigma^2}, \tag{11}$$

and $(x)^+$ denotes the larger of 0 and x.

B. Neither channel nor interference information at the transmitter

If the transmitter applies uniform power allocation across the transmit antennas, i.e., $\Sigma_s = (P_T/M)\mathbf{I}_M$, the capacity is given by

$$C = \log_2 \det\left(\mathbf{I}_N + \frac{P_T}{M \cdot \sigma^2}\mathbf{R}^{-1}\mathbf{H}\mathbf{H}^\dagger\right). \tag{12}$$

C. Only channel information at the transmitter

It is claimed in [3] that the optimal power allocation is the water-filling using $\mathbf{H}$ and assuming interference covariance matrix to be an identity matrix, i.e., the optimal transmit signal covariance matrix Σ_s is obtained from (8)-(11) by setting $\mathbf{R} = \mathbf{I}_N$, and the capacity is obtained by substituting the resultant Σ_s into (4). However, no justification that this scheme is optimal was given in [3]. At the same time, if we consider $\mathbf{R}^{-1/2}\mathbf{H}$ as a combined channel, without knowing $\mathbf{R}$, we do not know this combined channel. As a result, uniform power allocation at the transmitter should be used, i.e., the capacity is as (12). It is not obvious which power allocation scheme gives a higher

capacity, uniform power allocation (Section III-B) or water-filling (Section III-A) using $\mathbf{R} = \mathbf{I}_N$. Uniform power allocation does not use the known channel information, while the water-filling scheme uses the incorrect interference information. We simulated 10,000 sets of $\mathbf{H}$ and $\mathbf{R}$, and in all cases water-filling scheme using $\mathbf{H}$ only gives a higher capacity than uniform power allocation. However, no proof that this is true in general has been found.

IV. Channel Capacity with Estimated Channel and Interference

When the transmitter is provided with estimates of channel and/or interference covariance matrix, we can calculate the capacity by applying water-filling as in Section III-A using estimated interference covariance and channel matrices $\hat{\mathbf{R}}$ and $\hat{\mathbf{H}}$, respectively. As a result, we are able to evaluate the degradation of capacity due to estimation error of channel and interference covariance matrices.

We model the estimate of $\mathbf{H}$ as

$$\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E}_H \tag{13}$$

where $\mathbf{H}$ is the true channel. The elements in the estimation error matrix, $\mathbf{E}_H$, are i.i.d. zero-mean complex Gaussian. This implies that the estimation errors of channel are mutually independent. We assume that the variance of estimation error is proportional to the variance of true channel. Therefore, the variance of the (i,j)th element of $\mathbf{E}_H$, $\mathbf{E}_{H,ij}$, is specified by

$$\mathrm{VAR}(\mathbf{E}_{H,ij}) = \mu_H \cdot \mathrm{VAR}(\mathbf{H}_{ij}) \tag{14}$$

where μ_H is the parameter that controls the quality of the estimate. As the (i,j)th element in $\mathbf{H}$, $\mathbf{H}_{ij}$, is complex Gaussian with unit variance, $\mathrm{VAR}(\mathbf{H}_{ij}) = 1$.

Similarly, we model the estimate of $\mathbf{R}$ as

$$\hat{\mathbf{R}} = \mathbf{R} + \mathbf{E}_R \tag{15}$$

where $\mathbf{R}$ is the true interference covariance matrix. We restrict the estimation error matrix $\mathbf{E}_R$ to be Hermitian. We assume that the elements in the lower triangle of $\mathbf{E}_R$, $\mathbf{E}_{R,ij}$ for $i \leq j$, are mutually independent. The elements $\mathbf{E}_{R,ij}$ for $i < j$ are i.i.d. complex Gaussian, while the diagonal elements of $\mathbf{E}_R$ are i.i.d. real Gaussian. Again, the variance of $\mathbf{E}_{R,ij}$ is specified by

$$\mathrm{VAR}(\mathbf{E}_{R,ij}) = \mu_R \cdot \mathrm{VAR}(\mathbf{R}_{ij}). \tag{16}$$

The variance of the diagonal elements in $\mathbf{R}$ can be calculated as

$$\mathrm{VAR}(\mathbf{R}_{jj}) = \left(\frac{P_I}{\sigma^2 \cdot L}\right)^2 \sum_{i=1}^{L} \mathrm{VAR}(\mathbf{h}_{ij}\mathbf{h}_{ij}^{\dagger}) \tag{17}$$

where $\mathbf{h}_{ij}$ is the jth element in vector $\mathbf{h}_i$. Since $\mathbf{h}_{ij}$ is zero-mean complex Gaussian with unit variance, $\mathbf{h}_{ij}\mathbf{h}_{ij}^{\dagger}$ is chi-square distributed with 2 degree of freedom, and $\mathrm{VAR}(\mathbf{h}_{ij}\mathbf{h}_{ij}^{\dagger}) = 1$. As $\mathbf{h}_{ij}$'s are i.i.d. for all i and j, we have

$$\mathrm{VAR}(\mathbf{R}_{jj}) = \left(\frac{P_I}{\sigma^2}\right)^2 \frac{1}{L}. \tag{18}$$

The variance of off-diagonal elements in $\mathbf{R}$ is

$$\mathrm{VAR}(\mathbf{R}_{j_1 j_2}) = \left(\frac{P_I}{\sigma^2 \cdot L}\right)^2 \sum_{i=1}^{L} \mathrm{VAR}(\mathbf{h}_{ij_1}\mathbf{h}_{ij_2}^{\dagger}). \tag{19}$$

Let $\mathbf{h}_{ij_1} = a_1 + jb_1$, $\mathbf{h}_{ij_2} = a_2 + jb_2$, and a_1, a_2, b_1 and b_2 are i.i.d. zero-mean complex Gaussian with unit variance. It can be shown that $E(\mathbf{h}_{ij_1}\mathbf{h}_{ij_2}^{\dagger}) = 0$ and $\mathrm{VAR}(\mathbf{h}_{ij_1}\mathbf{h}_{ij_2}^{\dagger}) = 1$.

With specified μ_H and μ_R, we are able to simulate estimated channel and interference covariance matrices $\hat{\mathbf{H}}$ and $\hat{\mathbf{R}}$, respectively. The optimal transmit signal covariance matrix Σ_s is found by applying water-filling, i.e., (8)-(11) with estimates $\hat{\mathbf{H}}$ and/or $\hat{\mathbf{R}}$. The capacity is then obtained by substituting the resultant Σ_s into (4).

V. Simulation Results

We calculate the capacity under different assumptions of knowledge of channel and interference at the transmitter. For the case of only channel information at the transmitter, we use (8)-(11) and set $\mathbf{R} = \mathbf{I}_N$ to obtain the capacity. As $\mathbf{H}$ and $\mathbf{R}$ are random matrices, the capacity is treated as a random variable. The performance measurement here is the 10% outage capacity, $C_{0.1}$, where $P(C < C_{0.1}) = 10\%$. Monte Carlo simulation is used to obtain the 10% outage capacity.

In Fig. 1, we fix the total interference power and evaluate the outage capacity as the number of interferers increases. The user of interest is assumed to have 4 transmit and 4 receive antennas. The ratio of signal power to noise power and the ratio of total interference power to noise power are both 20dB. We find that the 10% outage capacity decreases significantly as the number of interferers increases. When the channel and interference are not known at the transmitter, the capacity with one interferer is 16 bps/Hz. This number is reduced sharply to 3 bps/Hz with 10 interferers each with one tenth the power. This implies that MIMO systems perform more efficiently where there are a few strong interferers.

In Fig. 2, we fix the number of interferers to be 4, and examine the outage capacity as we increase the ratio of total interference power to noise power. Again, the user of interest is assumed to have 4 transmit and 4 receive antennas. The ratio of signal power to noise power is 15dB. We observe that when the total interference power is low, knowing only the channel allows us to achieve about the same capacity as that in the case of full knowledge of channel and interference. However, when the total interference power is high, without interference information, knowing only the channel leads to about the same capacity as that in the case of no channel and interference knowledge at the transmitter.

In Fig. 3, assuming the user of interest has the same number of transmit and receive antennas, we calculate the outage capacity as the number of transmit antennas increases. We fix the number of interferers to be 4, the ratio of signal power to noise power and the ratio of total interference power to noise power are both 20dB. We observe that the capacity increases almost linearly as the number of antennas. In addition, the differences in capacity under different knowledge of channel and interference increase as the number of antennas increases.

In Fig. 4, we assess the degradation of channel capacity using estimated channel and/or interference. The user of interest is assumed to have 4 transmit and 4 receive antennas, and the ratio of signal power to noise power and the ratio of total interference power to noise power are both 20dB. The number of interferers is 10. In the case of no knowledge of channel and interference at the transmitter, we compare the capacity of uniform power allocation to that of water-filling using estimated channel and interference. We observe that, for $\mu_H = \mu_R$, when μ_H and μ_R are less than 50%, water-filling using estimated channel and interference achieves a higher capacity than uniform power allocation. In the case of only channel information at the transmitter, we compare the capacity of water-filling using known channel and estimated interference to that of water-filling using channel only. Again, for $\mu_H = \mu_R$, we observe that when μ_R is less than 50%, water-filling using known channel information and estimate of interference covariance matrix is better than water-filling using channel information only. Fig. 4 also shows the degradation of capacity due to estimation error of channel and interference for cases of $\mu_H = 0.1\mu_R$ and $\mu_H = 10\mu_R$.

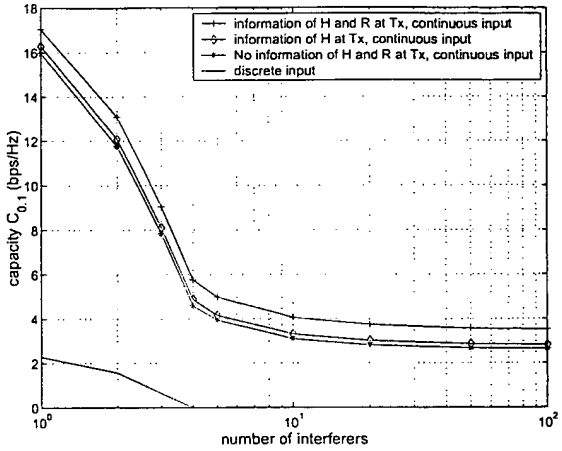


Fig. 1. 10% outage capacity versus number of interferers. The user of interest has 4 transmit and 4 receive antennas, the ratio of signal power to noise power is 20dB, and the ratio of total interference power to noise power is 20dB.

VI. Conclusions

In this paper, we fixed the total interference-plus-noise power and examined MIMO outage capacity under different interference environments: (1) a few high-data-rate interferers each with high power, (2) a large number of low-data-rate interferers each with low power. The results show that MIMO capacity is larger with fewer high-data-rate interferers. We also evaluated the degradation of outage capacity using estimated channel and/or interference.

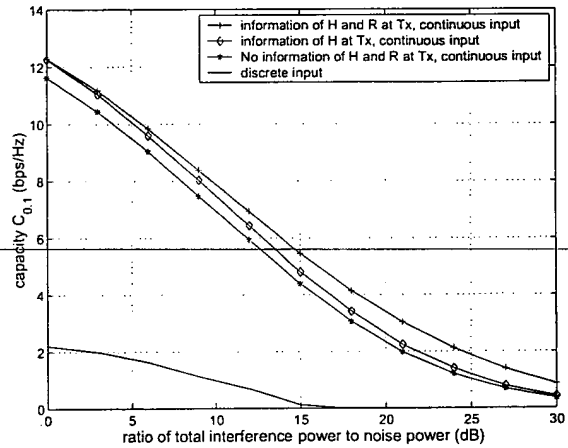


Fig. 2. 10% outage capacity versus the ratio of total interference power to noise power. The user of interest has 4 transmit and 4 receive antennas, the ratio of signal power to noise power is 15dB, and the number of interferers is 4.

References

[1] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 315-335, 1998.

[2] D. S. Shiu, *Wireless Communication Using Dual Antenna Arrays*. Kluwer Academic Publishers, 2000.

[3] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-Optimal Space-Time Processing with Multiple Transmit and Receive Antennas," *IEEE Commu. Letters*, pp. 85-87, March 2001.

[4] I. E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, pp. 585-595, Nov./Dec. 1999.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1990.

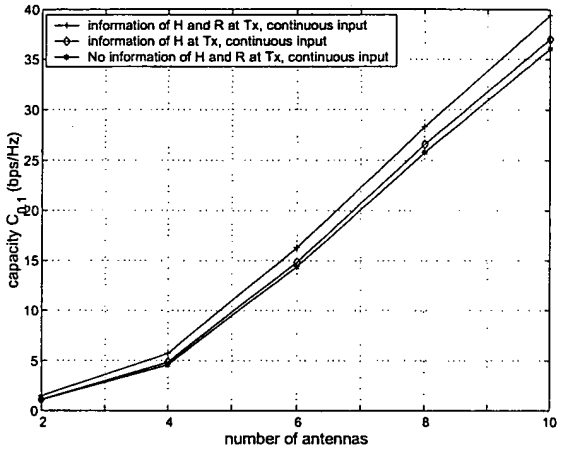


Fig. 3. 10% outage capacity versus number of antennas, assuming the user of interest has the same number of transmit and receive antennas. The number of interferers is 4, the ratio of signal power to noise power is 20dB, and the ratio of total interference power to noise power is 20dB.

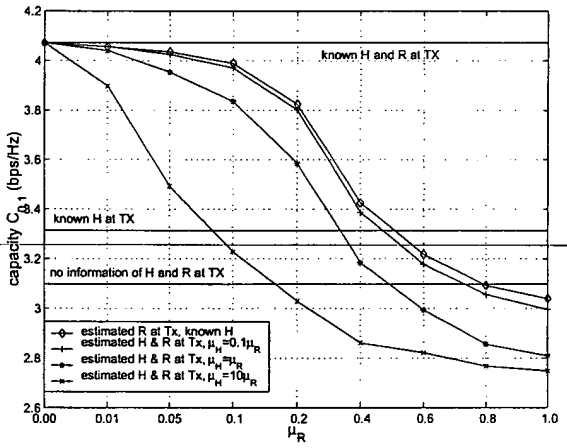


Fig. 4. 10% outage capacity versus μ_R. The user of interest has 4 transmit and 4 receive antennas, the number of interferers is 10, the ratio of signal power to noise power is 20dB, and the ratio of total interference power to noise power is 20dB.

Channel Estimation and Data Detection for MIMO Systems under Spatially and Temporally Colored Interference

Yi Song, *Student Member, IEEE*

Steven D. Blostein, *Senior Member, IEEE*

(Revised 2)

Yi Song and Steven D. Blostein are with the Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6, Tel: (613) 533-6561, E-mail: {songy, sdb}@ee.queensu.ca. The material in this paper was presented in part at IEEE VTC, Fall 2002.

Abstract

The impact of interference on multiple-input multiple-output (MIMO) systems has attracted recent interest. Most studies of channel estimation and data detection for MIMO systems consider spatially and temporally white interference at the receiver. In this paper, we address channel estimation, interference correlation estimation and data detection for MIMO systems under both spatially and temporally colored interference. We examine the case of one dominant interferer in which the data rate of the desired user could be the same or a multiple of that of the interferer. Assuming known temporal interference correlation as a benchmark, we derive maximum-likelihood estimates of the channel matrix and spatial interference correlation matrix, and apply these estimates to a generalized version of the BLAST (Bell Labs Layered Space-Time) ordered data detection algorithm. We then investigate the performance loss by not exploiting interference correlation. For a $(5, 5)$ MIMO system undergoing independent Rayleigh fading, we observe that exploiting both spatial and temporal interference correlation in channel estimation and data detection results in potential gains of 1.5dB and 4dB for an interferer operating at the same data rate and at half the data rate, respectively. Ignoring temporal correlation, it is found that spatial correlation accounts for about 1 db of this gain.

Index Terms

Multiple-input multiple-output, Interference, Channel estimation, Data detection

I. INTRODUCTION

Wireless systems with multiple transmitting and receiving antennas have been shown to have a large Shannon channel capacity in a rich scattering environment [1][2]. By transmitting parallel data streams over a multiple-input multi-output (MIMO) channel, it was shown that the Shannon capacity of the MIMO channel increases significantly with the number of transmitting and receiving antennas [2]. Layered space-time architectures were proposed for high-rate transmission in [3] and [4]. Space-time coding techniques have also been investigated [5][6].

While substantial research efforts have focussed on point-to-point MIMO link performance, the impact of interference on MIMO systems has received less interest. In a cellular environment, co-channel interference (CCI) from other cells exists due to channel reuse. In [7], channel capacities in the presence of spatially colored interference were derived under different assump-

tions of knowledge of the channel matrix and interference statistics at the transmitter. The impact of spatially colored interference on MIMO channel capacity was studied in [8][9][10]. The capacity of MIMO systems with interference in the limiting case of a large number of antennas was studied in [11]. The overall capacity of a group of users each employing a MIMO link was investigated in [12]. The output signal-to-interference-power ratio (SIR) was analytically calculated in [13] when a single data stream is transmitted over independent Rayleigh MIMO channels. While the majority of the studies deal with channel capacity, in this paper we will focus on the achievable symbol error rate performance of a MIMO link with interference.

Prior results on estimation of vector channels and spatial interference statistics for CDMA (code division multiple access) single-input multiple-output systems can be found in [14]. Most studies of channel estimation and data detection for MIMO systems assume spatially and temporally white interference. For example, in [15], maximum-likelihood (ML) estimation of the channel matrix using training sequences was presented assuming temporally white interference. Assuming perfect knowledge of the channel matrix at the receiver, ordered zero-forcing (ZF) and minimum mean-squared error (MMSE) detection were studied for both spatially and temporally white interference in [4] and [16], respectively. However, in cellular systems, the interference is, in general, both spatially and temporally colored.

In this paper, we propose and study a new algorithm that jointly estimates the channel matrix and the spatial interference correlation matrix in a maximum-likelihood framework. We develop a multi-vector-symbol MMSE data detector that exploits interference correlation. In the case of a single dominant interferer and large signal-to-noise ratio (SNR), we show that spatial and temporal second-order interference statistics can be decoupled in the form of a matrix Kronecker product. In finite SNR, the decoupling of spatial and temporal statistics of interference-plus-noise is only an approximation. We also determine the conditions where this approximation breaks down.

Although temporal interference correlation is difficult to estimate in practice, our objectives are to determine the performance benchmark achieved if temporal correlation were known. As

sources of temporal correlation, we consider cases in which the data rate of the desired user is either the same as or a multiple of that of the interferer. The new ML algorithm serves as a performance benchmark when temporal and spatial interference correlation are exploited in joint channel estimation and data detection. We also assess the performance improvement obtained in more practical cases where only part of the correlation information is exploited, including the performance obtained by assuming temporally white interference, i.e., ignoring temporal correlation.

The paper is organized as follows. In Section II, we present our system model of temporal and spatial interference. In Section III, we derive ML estimates of channel and spatial interference correlation matrices assuming known temporal interference correlation. In Section IV, one-vector-symbol detection is extended to a multi-vector-symbol version which is used to exploit temporal interference correlation. In Section V, we consider the case of one interferer and large SNR and assess the benefits of taking temporal and/or spatial interference correlation into account for channel estimation and data detection. We then examine the level of SNR at which the approximation of separate spatial and temporal interference-plus-noise statistics break down. In cases where the spatial and temporal correlation are not separable, the performance improvement obtained by exploiting the spatial correlation is evaluated. For reference, comparisons are made to the well-known Direct Matrix Inversion (DMI) algorithm [25], generalized to multiple input signals, a batch method that does not require estimates of channel and spatial interference correlation matrices.

In the paper, notation $(\cdot)^T$ refers to transpose, $(\cdot)^*$ refers to conjugate, $(\cdot)^\dagger$ refers to conjugate transpose, and $\boldsymbol{I}_N$ refers to an $N \times N$ identity matrix.

II. System Model

We consider a single-user link consisting of N_t transmitting and N_r receiving antennas, denoted as (N_t, N_r). The desired user transmits data frame by frame. Each frame has M data vectors. The first N data vectors are used for training so that the desired user's channel matrix

and interference statistics can be estimated, and the remaining data vectors are for information transmission. In a slow flat fading environment, the received signal vector at time j is expressed as

$$\boldsymbol{y}_j = \boldsymbol{H}\boldsymbol{x}_j + \boldsymbol{n}_j, \quad j = 0, \cdots, M-1 \tag{1}$$

where $\boldsymbol{x}_j$ is the transmitted data vector, $\boldsymbol{H}$ is the $N_r \times N_t$ spatial channel gain matrix, and interference vector $\boldsymbol{n}_j$ is zero-mean circularly symmetric complex Gaussian. We assume channel matrix $\boldsymbol{H}$ is fixed during one frame. This is a reasonable assumption since high speed data services envisioned for MIMO systems are generally intended for low mobility users. By the same argument, it is also assumed that the interference statistics are fixed during one frame.

In practice, the interference may be both spatially and temporally correlated. We assume that the cross-correlation between the interference vectors at time i and j is $E\left\{\boldsymbol{n}_i \boldsymbol{n}_j^\dagger\right\} = \Lambda_M(i,j)\boldsymbol{R}$ where $\Lambda_M(i,j)$ is the (i,j)th element of an $M \times M$ matrix Λ_M. The (i,j)th element of matrix $\boldsymbol{R}$ is the correlation between the ith and jth elements of interference vector $\boldsymbol{n}_k$, $k \in 0, \cdots, M-1$. As a result, the covariance matrix of the concatenated interference vector $\bar{\boldsymbol{n}} = [\boldsymbol{n}_0^T \cdots \boldsymbol{n}_{M-1}^T]^T$ is

$$E\left\{\bar{\boldsymbol{n}}\bar{\boldsymbol{n}}^\dagger\right\} = \begin{bmatrix} \Lambda_M(0,0)\boldsymbol{R} & \cdots & \Lambda_M(0,M-1)\boldsymbol{R} \\ \vdots & & \vdots \\ \Lambda_M(M-1,0)\boldsymbol{R} & \cdots & \Lambda_M(M-1,M-1)\boldsymbol{R} \end{bmatrix} = \Lambda_M \otimes \boldsymbol{R} \tag{2}$$

where $\otimes$ denotes Kronecker product, and matrices Λ_M and $\boldsymbol{R}$ capture the temporal and spatial correlation of the interference, respectively. The above model implies that the spatial and temporal interference statistics are separable. The correlation matrices Λ_M and $\boldsymbol{R}$ are determined by the application-specific signal model. In Section V, we provide an example in which the interference covariance matrix has the above Kronecker product form. When the interference statistics can only be approximated by (2), the conditions where this approximation breaks down are investigated in Section V-D.3. In addition to interference correlation, we remark that a decoupled temporal and spatial correlation structure arises in the statistics of fading vector channels consisting of a mobile with one antenna and a base station with an antenna array [17].

III. Joint Estimation of Channel and Spatial Interference Statistics

During a training period of N vector symbols, we concatenate the received signal vectors, the training signal vectors and the interference vectors as $\bar{\boldsymbol{y}} = [\boldsymbol{y}_0^T \cdots \boldsymbol{y}_{N-1}^T]^T$, $\bar{\boldsymbol{x}} = [\boldsymbol{x}_0^T \cdots \boldsymbol{x}_{N-1}^T]^T$ and $\bar{\boldsymbol{n}} = [\boldsymbol{n}_0^T \cdots \boldsymbol{n}_{N-1}^T]^T$, respectively. The received signal in (1) is rewritten as the vector

$$\bar{\boldsymbol{y}} = (\boldsymbol{I}_N \otimes \boldsymbol{H})\bar{\boldsymbol{x}} + \bar{\boldsymbol{n}}$$

where $\bar{\boldsymbol{n}}$ is circularly symmetric complex Gaussian with zero-mean and covariance matrix $\Lambda_N \otimes \boldsymbol{R}$. Assuming prior knowledge of temporal interference correlation matrix Λ_N, we need to estimate channel matrix $\boldsymbol{H}$ and spatial interference correlation matrix $\boldsymbol{R}$. If $\boldsymbol{R}$ and Λ_N are nonsingular, the conditional probability density function (PDF)

$$\Pr(\bar{\boldsymbol{y}}|\boldsymbol{H},\boldsymbol{R}) = \frac{1}{\pi^{N \cdot N_r} \det(\Lambda_N \otimes \boldsymbol{R})} \exp\left\{ -[\bar{\boldsymbol{y}} - (\boldsymbol{I}_N \otimes \boldsymbol{H})\bar{\boldsymbol{x}}]^\dagger (\Lambda_N \otimes R)^{-1} [\bar{\boldsymbol{y}} - (\boldsymbol{I}_N \otimes \boldsymbol{H})\bar{\boldsymbol{x}}] \right\}. \tag{3}$$

A. Maximum-likelihood solution

The ML estimate of the pair of matrices $(\boldsymbol{H}, \boldsymbol{R})$ is the value of (H, R) that maximizes the conditional PDF in (3), which is equivalent to maximizing $\ln \Pr(\bar{\boldsymbol{y}}|\boldsymbol{H}, \boldsymbol{R})$.

Letting $\boldsymbol{A}$ and $\boldsymbol{B}$ denote $m \times m$ and $n \times n$ square matrices, and using identities [18]

$$\det(\boldsymbol{A} \otimes \boldsymbol{B}) = \det(\boldsymbol{A})^n \det(\boldsymbol{B})^m$$

and

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}, \quad \boldsymbol{A} \text{ and } \boldsymbol{B} \text{ are nonsingular,}$$

it can be shown that maximizing (3) is equivalent to minimizing

$$f(\boldsymbol{H}, \boldsymbol{R}) = \ln \det(\boldsymbol{R}) + \frac{1}{N} [\bar{\boldsymbol{y}} - (\boldsymbol{I}_N \otimes \boldsymbol{H})\bar{\boldsymbol{x}}]^\dagger (\Lambda_N^{-1} \otimes \boldsymbol{R}^{-1}) [\bar{\boldsymbol{y}} - (\boldsymbol{I}_N \otimes \boldsymbol{H})\bar{\boldsymbol{x}}]. \tag{4}$$

Denoting the elements of Λ_N^{-1} as

$$\Lambda_N^{-1} = \begin{bmatrix} \alpha_{0,0} & \cdots & \alpha_{0,N-1} \\ \vdots & & \vdots \\ \alpha_{N-1,0} & \cdots & \alpha_{N-1,N-1} \end{bmatrix}, \tag{5}$$

we rewrite (4) as

$$f(\boldsymbol{H},\boldsymbol{R}) = \ln\det(\boldsymbol{R}) + \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}\left(\boldsymbol{y}_i-\boldsymbol{H}\boldsymbol{x}_i\right)^{\dagger}\boldsymbol{R}^{-1}\left(\boldsymbol{y}_j-\boldsymbol{H}\boldsymbol{x}_j\right)$$

$$= \ln\det(\boldsymbol{R}) + \text{trace}\left\{\boldsymbol{R}^{-1}\frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}\left(\boldsymbol{y}_i-\boldsymbol{H}\boldsymbol{x}_i\right)\left(\boldsymbol{y}_j-\boldsymbol{H}\boldsymbol{x}_j\right)^{\dagger}\right\}. \quad (6)$$

To find the value of $(\boldsymbol{H},\boldsymbol{R})$ that minimizes $f(\boldsymbol{H},\boldsymbol{R})$ in (6), we set $\partial f(\boldsymbol{H},\boldsymbol{R})/\partial\boldsymbol{H} = \boldsymbol{0}$. Define the weighted sample correlation matrices[1] as

$$\tilde{\boldsymbol{R}}_{yy} = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}\boldsymbol{y}_i\boldsymbol{y}_j^{\dagger}, \quad (7)$$

$$\tilde{\boldsymbol{R}}_{xy} = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}\boldsymbol{x}_i\boldsymbol{y}_j^{\dagger}, \quad (8)$$

and

$$\tilde{\boldsymbol{R}}_{xx} = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}\boldsymbol{x}_i\boldsymbol{x}_j^{\dagger}. \quad (9)$$

Using identities of matrix derivative [18], it can be shown [19] that (6) is minimized by

$$\hat{\boldsymbol{H}} = \tilde{\boldsymbol{R}}_{xy}^{\dagger}\tilde{\boldsymbol{R}}_{xx}^{-1}. \quad (10)$$

Setting $\partial f(\hat{\boldsymbol{H}},\boldsymbol{R})/\partial\boldsymbol{R} = \boldsymbol{0}$, it can also be shown that the estimate of spatial interference correlation matrix is given by

$$\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}\left(\boldsymbol{y}_i-\hat{\boldsymbol{H}}\boldsymbol{x}_i\right)\left(\boldsymbol{y}_j-\hat{\boldsymbol{H}}\boldsymbol{x}_j\right)^{\dagger} \quad (11)$$

$$= \tilde{\boldsymbol{R}}_{yy} - \hat{\boldsymbol{H}}\tilde{\boldsymbol{R}}_{xy}. \quad (12)$$

We remark that if $\tilde{\boldsymbol{R}}_{xy}$ and $\tilde{\boldsymbol{R}}_{xx}$ in (10) were instead known cross- and auto-correlation matrices, the estimate for $\boldsymbol{H}$ would represent the Wiener solution.

[1]To distinguish weighted sample correlation matrices from conventional sample correlation matrices in Section III-B, we denote the former by a tilde and the latter without a tilde.

B. Special case: temporally white interference

If interference is temporally white, with loss of generality, we may substitute $\Lambda_N = \boldsymbol{I}_N$ into (7)-(12), and obtain estimates

$$\hat{\boldsymbol{H}}_w = \boldsymbol{R}_{xy}^{\dagger} \boldsymbol{R}_{xx}^{-1}, \tag{13}$$

and

$$\hat{\boldsymbol{R}}_w = \boldsymbol{R}_{yy} - \hat{\boldsymbol{H}}_w \boldsymbol{R}_{xy} \tag{14}$$

where the subscript w indicates temporally white interference, and the sample correlation matrices

$$\boldsymbol{R}_{yy} = \frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{y}_i \boldsymbol{y}_i^{\dagger}, \tag{15}$$

$$\boldsymbol{R}_{xy} = \frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{x}_i \boldsymbol{y}_i^{\dagger}, \tag{16}$$

and

$$\boldsymbol{R}_{xx} = \frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{x}_i \boldsymbol{x}_i^{\dagger}. \tag{17}$$

Note that $\hat{\boldsymbol{H}}_w$ in (13) is the same as the channel estimate used in [15].

C. Whitening filter interpretation

To obtain insight on the estimates in (10) and (12), we let the received signal vectors during the training period undergo a linear transformation where the transformed received signal vectors are

$$[\boldsymbol{y}'_0 \dots \boldsymbol{y}'_{N-1}] = [\boldsymbol{y}_0 \dots \boldsymbol{y}_{N-1}] \Lambda_N^{-1/2}.$$

At the output of the transformation, we have

$$\boldsymbol{y}'_i = \boldsymbol{H} \boldsymbol{x}'_i + \boldsymbol{n}'_i, \quad i = 0, \dots, N-1,$$

where the transformed training signal vectors and interference vectors are

$$[\boldsymbol{x}'_0 \dots \boldsymbol{x}'_{N-1}] = [\boldsymbol{x}_0 \dots \boldsymbol{x}_{N-1}] \Lambda_N^{-1/2} \quad \text{and} \quad [\boldsymbol{n}'_0 \dots \boldsymbol{n}'_{N-1}] = [\boldsymbol{n}_0 \dots \boldsymbol{n}_{N-1}] \Lambda_N^{-1/2},$$

respectively. Concatenating the transformed interference vectors as $\bar{n}' = [n_0'^T \dots n_{N-1}'^T]^T$, it can be shown that

$$\bar{n}' = (\Lambda_N^{-1/2} \otimes I_{N_r})\bar{n}$$

where $\bar{n} = [n_0^T \dots n_{N-1}^T]^T$. Since the covariance matrix of $\bar{n}$ is $\Lambda_N \otimes R$, the covariance matrix of $\bar{n}'$ is

$$\begin{aligned} \text{cov}(\bar{n}') &= (\Lambda_N^{-1/2} \otimes I_{N_r})\text{cov}(\bar{n})(\Lambda_N^{-1/2} \otimes I_{N_r})^\dagger \\ &= (\Lambda_N^{-1/2} \otimes I_{N_r})(\Lambda_N \otimes R)(\Lambda_N^{-1/2} \otimes I_{N_r}) \\ &= I_N \otimes R \end{aligned} \tag{18}$$

where we used $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$ and $(A \otimes B)(C \otimes D) = AC \otimes BD$ [18]. We also used the fact that temporal correlation matrix Λ_N is symmetric, as well as $\Lambda_N^{-1/2}$. From (18), it is obvious that the transformed interference vectors $\{n_0' \dots n_{N-1}'\}$ are temporally white with spatial correlation matrix R.

As a result, we can estimate H and R from the sample correlation matrices of transformed signal vectors as in Section III-B. The sample correlation matrix

$$\begin{aligned} R_{y'y'} &= \frac{1}{N}\sum_{i=0}^{N-1} y_i' y_i'^\dagger \\ &= \frac{1}{N}[y_0' \dots y_{N-1}'][y_0' \dots y_{N-1}']^\dagger \\ &= \frac{1}{N}[y_0 \dots y_{N-1}]\Lambda_N^{-1/2}\Lambda_N^{-\dagger/2}[y_0 \dots y_{N-1}]^\dagger \\ &= \frac{1}{N}[y_0 \dots y_{N-1}]\Lambda_N^{-1}[y_0 \dots y_{N-1}]^\dagger = \tilde{R}_{yy}, \end{aligned}$$

which shows that the weighted sample correlation matrix of $\{y_0 \dots y_{N-1}\}$ is equivalent to the sample correlation matrix of $\{y_0' \dots y_{N-1}'\}$. Similarly, the weighted sample correlation matrices $\tilde{R}_{xy}$ and $\tilde{R}_{xx}$ are equivalent to the sample correlation matrices $R_{x'y'}$ and $R_{x'x'}$, respectively. Therefore, the estimates in (10) and (12) can also be realized by first temporally whitening the interference, and then forming the estimates from the sample correlation matrices of the transformed signal vectors.

IV. Data Detection

We focus on ordered MMSE detection due to the better performance of MMSE compared to ZF detection [20]. For received signal vector $\boldsymbol{y}_i = \boldsymbol{H}\boldsymbol{x}_i + \boldsymbol{n}_i$, modifying the BLAST algorithm in [16], the steps of ordered MMSE detection of $\boldsymbol{x}_i$ from $\boldsymbol{y}_i$ with estimated channel and interference spatial correlation matrices are as follows:

Step 1 Initialization: set $k = 1$, $\boldsymbol{H}_k = \hat{\boldsymbol{H}}$, $\tilde{\boldsymbol{x}}_k = \boldsymbol{x}_i$, $\tilde{\boldsymbol{y}}_k = \boldsymbol{y}_i$.

Step 2 Calculate the estimation error covariance matrix $\boldsymbol{P}_k = (\boldsymbol{I}_{N_t+1-k} + \boldsymbol{H}_k^{\dagger}\hat{\boldsymbol{R}}^{-1}\boldsymbol{H}_k)^{-1}$. Find $m = \arg\min_j \boldsymbol{P}_k(j,j)$ where $\boldsymbol{P}_k(j,j)$ denotes the jth diagonal element of $\boldsymbol{P}_k$. Hence, the mth signal component of $\tilde{\boldsymbol{x}}_k$ has the smallest estimation error variance.

Step 3 Calculate the weighting matrix $\boldsymbol{A}_k = (\boldsymbol{I}_{N_t+1-k} + \boldsymbol{H}_k^{\dagger}\hat{\boldsymbol{R}}^{-1}\boldsymbol{H}_k)^{-1}\boldsymbol{H}_k^{\dagger}\hat{\boldsymbol{R}}^{-1}$. The mth element of $\tilde{\boldsymbol{x}}_k$ is estimated by $\hat{x}_k^m = Q\left(\boldsymbol{A}_k(m,:)\,\tilde{\boldsymbol{y}}_k\right)$ where $\boldsymbol{A}_k(m,:)$ denotes the mth row of matrix $\boldsymbol{A}_k$ and $Q(\cdot)$ denotes the slicing operation appropriate to the signal constellation.

Step 4 Assuming the detected signal is correct, remove the detected signal from the received signal, $\tilde{\boldsymbol{y}}_{k+1} = \tilde{\boldsymbol{y}}_k - \hat{x}_k^m \boldsymbol{H}_k(:,m)$ where $\boldsymbol{H}_k(:,m)$ denotes the mth column of $\boldsymbol{H}_k$.

Step 5 $\boldsymbol{H}_{k+1}$ is obtained by eliminating the mth column of matrix $\boldsymbol{H}_k$. $\tilde{\boldsymbol{x}}_{k+1}$ is obtained by eliminating the mth component of vector $\tilde{\boldsymbol{x}}_k$.

Step 6 If $k < N_t$, increment k and go to Step 2.

We refer this scheme as *one-vector-symbol detection* as we detect $\boldsymbol{x}_i$ using $\boldsymbol{y}_i$ only.

~~When interference is temporally colored, there may be performance to be gained by taking~~ the temporal interference correlation into account. That is, we may use $\boldsymbol{y}_{N+1}, \ldots, \boldsymbol{y}_M$ to detect $\boldsymbol{x}_{N+1}, \ldots, \boldsymbol{x}_M$ jointly where N is the training length and M is the frame length. Due to the complexity of using all the received signal vectors and for simplicity of presentation, we consider *two-vector-symbol detection* in which $(\boldsymbol{y}_i, \boldsymbol{y}_{i+1})$ is used to detect (x_i, x_{i+1}) jointly. The one-

vector-symbol algorithm can be easily extended to the two-vector-symbol version by writing

$$\underbrace{\begin{bmatrix} y_i \\ y_{i+1} \end{bmatrix}}_{\breve{y}_i} = \underbrace{\begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix}}_{\breve{H}} \underbrace{\begin{bmatrix} x_i \\ x_{i+1} \end{bmatrix}}_{\breve{x}_i} + \underbrace{\begin{bmatrix} n_i \\ n_{i+1} \end{bmatrix}}_{\breve{n}_i}.$$

With the estimated channel, an estimate of $\breve{H}$, denoted as $\hat{\breve{H}}$, can be obtained. Using the estimated spatial interference correlation and the known temporal interference correlation, we are able to estimate the covariance matrix of $\breve{n}_i$, denoted as $\hat{\breve{R}}$. Replacing x_i, y_i, $\hat{H}$ and $\hat{R}$ in the one-vector-symbol algorithm by $\breve{x}_i$, $\breve{y}_i$, $\hat{\breve{H}}$ and $\hat{\breve{R}}$, respectively, we obtain the two-vector-symbol detection algorithm.

V. Applications

In this section, we apply the channel estimation in Section III and data detection in Section IV to the case of a single-user link with one dominant co-channel interferer operating at different data rates.

A. System model

Consider a desired user with one dominant co-channel interferer. The assumption of one co-channel interferer can apply to cellular TDMA or FDMA systems when sectoring is used. For example, in 7-cell reuse systems, with 60 degree sectors, the number of co-channel interfering cells would be reduced to one [21]. We assume that desired and interfering users have N_t and L transmitting antennas, respectively, and that there are N_r receiving antennas. Assuming thermal noise is small relative to interference, we ignore thermal noise in the problem formulation. An investigation of this assumption in channels with noise appears in Section V-D.3. We also assume that over the duration of a transmitted frame, a randomly delayed replica of the interfering signal is transmitted continuously and that the interference statistics do not change. This assumption may not hold for asynchronous packet transmission systems. In a slow flat fading

environment, the vector signal at receiving antennas is

$$\boldsymbol{y}(t) = \sqrt{\frac{P_s T}{N_t}} \boldsymbol{H} \sum_{k=0}^{M-1} \boldsymbol{x}_k \tilde{g}(t - kT) + \sqrt{\frac{P_I T_I}{L}} \boldsymbol{H}_I \sum_{k=-\infty}^{\infty} b_k \tilde{g}_I(t - kT_I - \tau) \qquad (19)$$

where M is the frame length, and $\boldsymbol{H}$ ($N_r \times N_t$) and $\boldsymbol{H}_I$ ($N_r \times L$) are the channel matrices of the desired and interfering users, respectively. The channel matrices are also assumed fixed over a frame, and have independent realizations from frame to frame. The data transmission rates of the desired and interfering users are $1/T$ and $1/T_I$, respectively. The spectra of transmit impulse responses $\tilde{g}(t)$ and $\tilde{g}_I(t)$ are square-root raised cosines with parameters T and T_I, respectively. The same rolloff factor, β, is assumed for both $\tilde{g}(t)$ and $\tilde{g}_I(t)$. The data vectors of the desired and interfering users are $\boldsymbol{x}_k$ ($N_t \times 1$) and $\boldsymbol{b}_k$ ($L \times 1$), respectively. We assume that data symbols in $\boldsymbol{x}_k$'s and $\boldsymbol{b}_k$'s are mutually independent, zero-mean and with unit variance. We denote P_s and P_I as the transmit powers of the desired and interfering users, respectively. The delay of the interfering user relative to the desired user is τ, assumed to lie in $0 \leq \tau < T_I$.

Passing $\boldsymbol{y}(t)$ in (19) through a filter matched to the transmit impulse response of the desired user, $\tilde{g}(t)$, the vector signal at the output of the matched filter is

$$\boldsymbol{y}_{\text{MF}}(t) = \sqrt{\frac{P_s T}{N_t}} \boldsymbol{H} \sum_{k=0}^{M-1} \boldsymbol{x}_k g(t - kT) + \sqrt{\frac{P_I T_I}{L}} \boldsymbol{H}_I \sum_{k=-\infty}^{\infty} b_k g_I(t - kT_I - \tau) \qquad (20)$$

where $g(t) = \tilde{g}(t) * \tilde{g}(t)$, $g_I(t) = \tilde{g}_I(t) * \tilde{g}(t)$, and $*$ denotes convolution. As a result, $g(t)$ has a raised cosine spectrum and satisfies the Nyquist condition for zero intersymbol interference.

Assuming perfect synchronization for the desired user, as we sample the output of the matched filter (20) at time $t = jT$, we obtain

$$\boldsymbol{y}_j = \sqrt{\frac{P_s T}{N_t}} \boldsymbol{H} \boldsymbol{x}_j + \underbrace{\sqrt{\frac{P_I T_I}{L}} \boldsymbol{H}_I \sum_{k=-\infty}^{\infty} \boldsymbol{b}_k g_I(jT - kT_I - \tau)}_{\boldsymbol{n}_j}. \qquad (21)$$

The interference vector $\boldsymbol{n}_j$ is zero-mean since the data vector of interferer $\boldsymbol{b}_k$ is zero-mean. Note that there is no intersymbol interference for the desired user. However, due to the interferer's delay and/or mismatch between the transmit and receive impulse responses, intersymbol interference exists for the interferer.

B. Interference statistics

The cross-correlation between the interference vectors in (21) at time jT and qT is

$$\begin{aligned} & E\left\{\boldsymbol{n}_j \boldsymbol{n}_q^\dagger\right\} \\ = & \frac{P_I T_I}{L} \boldsymbol{H}_I \cdot E\left\{\left(\sum_{k_1=-\infty}^{\infty} \boldsymbol{b}_{k_1} g_I(jT - k_1 T_I - \tau)\right)\left(\sum_{k_2=-\infty}^{\infty} \boldsymbol{b}_{k_2}^\dagger g_I(qT - k_2 T_I - \tau)\right)\right\} \boldsymbol{H}_I^\dagger \\ = & \frac{P_I T_I}{L} \boldsymbol{H}_I \boldsymbol{H}_I^\dagger \cdot \sum_{k=-\infty}^{\infty} \left\{g_I(jT - kT_I - \tau) g_I(qT - kT_I - \tau)\right\}, \end{aligned}$$

where the last equality is due to the facts that $E\left\{\boldsymbol{b}_{k_1} \boldsymbol{b}_{k_2}^\dagger\right\} = 0$ for $k_1 \neq k_2$ and $E\left\{\boldsymbol{b}_k \boldsymbol{b}_k^\dagger\right\} = \boldsymbol{I}_L$.

During a training period of N vector symbols, the covariance matrix of the concatenated interference vector $\bar{\boldsymbol{n}} = [\boldsymbol{n}_0^T \cdots \boldsymbol{n}_{N-1}^T]^T$ has the form of (2) where

$$\Lambda_N(j,q) = \sum_{k=-\infty}^{\infty} \left\{g_I(jT - kT_I - \tau) g_I(qT - kT_I - \tau)\right\}, \quad 0 \leq j,q \leq N-1 \tag{22}$$

and

$$\boldsymbol{R} = \frac{P_I T_I}{L} \boldsymbol{H}_I \boldsymbol{H}_I^\dagger. \tag{23}$$

The $N_r \times N_r$ spatial correlation matrix $\boldsymbol{R}$ is determined by the interferer's channel matrix. The $N \times N$ temporal correlation matrix Λ_N depends on parameters T and T_I, delay τ and pulse $g_I(t)$; it can be calculated a priori if these parameters are known. The temporal correlation is due to intersymbol interference in the sampled interfering signal. We remark that for the case of multiple interferers with the same delay, the covariance matrix of interference also has the form of (2).

We study temporal interference correlation in the cases where (1) the interferer has the same data rate as that of the desired signal ($T = T_I$), and (2) the data rate of the desired user is an integer multiple of that of the interferer ($T_I = mT$, $m > 1$).

1) Interferer at the same data rate as desired signal: With $T = T_I$, $g_I(t)$ has a raised cosine spectrum and is given by [22]

$$g_I(t) = \text{sinc}(\pi t/T) \frac{\cos(\pi \beta t/T)}{1 - 4\beta^2 t^2/T^2}.$$

We note that $\mathbf{\Lambda}_N(j,q)$ depends on $j-q$. This indicates that the sequence consisting of interference vectors is stationary. Hence, the temporal correlation matrix is symmetric Toeplitz. By appropriate truncation of the infinite series in (22), we can numerically calculate the temporal correlation matrix. For the case of $\beta = 1$, $T = 1$ and $\tau = 0.5$, the elements of the temporal correlation matrix are

$$\mathbf{\Lambda}_N(j,q) = \begin{cases} 0.5 & j = q \\ 0.25 & |j-q| = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } 0 \le j,q \le N-1. \tag{24}$$

2) Interferer at a lower data rate than desired signal: It can be shown that $g_I(t)$ is given by

$$g_I(t) = \mathcal{F}^{-1}\left\{\sqrt{G_{rc,T_I}(f)}\sqrt{G_{rc,T}(f)}\right\}$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform, and $G_{rc,T}(f)$ is the raised cosine Fourier spectrum with parameter T and rolloff factor β. Unlike in the case of same-data-rate interferer where $\mathbf{\Lambda}_N(j,q)$ depends on $j-q$, in the case of lower-data-rate interferer, $\mathbf{\Lambda}_N(j,q)$ depends on the values of j and q. This indicates that the sequence consisting of interference vectors is cyclostationary [22][23]. It can be shown that $\mathbf{\Lambda}_N(j,q)$ is periodic with period m, i.e., $\mathbf{\Lambda}_N(j,q) = \mathbf{\Lambda}_N(j+m,q+m)$. As a result, the temporal correlation matrix $\mathbf{\Lambda}_N$ is symmetric, but not Toeplitz. Furthermore, for $N \ge m$, the number of nontrivial eigenvalues of $\mathbf{\Lambda}_N$ is $\lceil N/m \rceil$ where $\lceil \cdot \rceil$ rounds the argument to the nearest integer towards infinity [24]. For the case of $T_I = 2T$, $T = 1$, $\beta = 1$, $\tau = 0.25$ and training length $N = 6$, by numerical calculation of (22) with appropriate series truncation, the temporal correlation matrix is

$$\mathbf{\Lambda}_6 = \begin{bmatrix} 0.648 & 0.400 & -0.048 & -0.006 & -0.010 & -0.001 \\ 0.400 & 0.277 & 0.105 & 0.084 & 0.002 & 0.011 \\ -0.048 & 0.105 & 0.648 & 0.400 & -0.048 & -0.006 \\ -0.006 & 0.084 & 0.400 & 0.277 & 0.105 & 0.084 \\ -0.010 & 0.002 & -0.048 & 0.105 & 0.648 & 0.400 \\ -0.001 & 0.011 & -0.006 & 0.084 & 0.400 & 0.277 \end{bmatrix}. \tag{25}$$

Note that Λ_6 in (25) is singular as the number of nontrivial eigenvalues is 3.

C. Data detection without estimating channel and interference

During a training period of N symbol vectors, instead of estimating the channel matrix and interference statistics, one can alternatively employ a least-squares (LS) estimate of matrix $\boldsymbol{M}$ which minimizes the average estimation error

$$f_2(\boldsymbol{M}) = \text{trace}\left\{\frac{1}{N}\sum_{i=0}^{N-1}\left(\boldsymbol{x}_i - \boldsymbol{M}\boldsymbol{y}_i\right)\left(\boldsymbol{x}_i - \boldsymbol{M}\boldsymbol{y}_i\right)^{\dagger}\right\}.$$

By setting $\partial f_2(\boldsymbol{M})/\partial \boldsymbol{M} = \boldsymbol{0}$, we obtain

$$\boldsymbol{M} = \boldsymbol{R}_{xy}\boldsymbol{R}_{yy}^{-1} \tag{26}$$

where sample correlation matrices $\boldsymbol{R}_{xy}$ and $\boldsymbol{R}_{yy}$ are defined in (16) and (15), respectively. The transmitted signal vector $\boldsymbol{x}_i$ is detected as $Q\left(\boldsymbol{M}\boldsymbol{y}_i\right)$ where $Q(\cdot)$ is the slicing operation appropriate to the signal constellation. We remark that (26) is the well-known Direct Matrix Inversion (DMI) algorithm [25] generalized for multiple input signals. A significant loss in performance is expected for this LS detector since without estimates of channel and spatial interference correlation matrices, iterative MMSE detection cannot be performed.

D. Simulation results

Monte-Carlo simulations are used to assess the benefits of taking temporal and spatial interference correlation into account for channel estimation and data detection in the case of one interferer. Although temporal interference correlation may be difficult to estimate in practice, we examine this as a benchmark and determine the performance loss due to ignoring this correlation. We evaluate average symbol error rates (SERs) in independent Rayleigh fading channels of rich scattering, i.e., the elements in channel matrices $\boldsymbol{H}$ and $\boldsymbol{H}_I$ are independent, identically distributed (i.i.d.) zero-mean complex Gaussian with unit variance. We assume that the

desired user has 5 transmitting and 5 receiving antennas, and the interfering user has 6 transmitting antennas[2]. Both the desired and interfering users employ uncoded QPSK modulation. The training signal vectors are columns of an FFT matrix [16] to guarantee orthogonal training sequences from different transmitting antennas. We define signal-to-interference-power-ratio SIR(dB) $= 10 \log P_s/P_I$. Without loss of generality, we set $P_I = 1$ in the simulation. The SERs of two cases are simulated: (1) interferer at the same data rate as the desired signal, and (2) the data rate of the desired user is twice that of the interferer.

In Figs. 1 to 4, with solid and dashed lines representing one- and two-vector-symbol data detection, respectively, we plot average SERs for the following cases:

(a) perfectly known channel parameters and interference statistics, with one-vector-symbol (curve 1) and two-vector-symbol (curve 2) detection;

(b) channel and spatial interference correlation matrices are estimated assuming known temporal interference correlation, with one-vector-symbol (curve 3) and two-vector-symbol (curve 4) detection;

(c) channel and spatial interference correlation matrices are estimated assuming temporally white interference, with one-vector-symbol detection (curve 5);

(d) only the channel matrix $\boldsymbol{H}$ is estimated assuming temporally white interference; an identity spatial interference correlation matrix is used in one-vector-symbol data detection (curve 6).

(e) LS estimate of the transmitted signal vector without ordered detection (Section V-C) (curve 7).

We remark that cases (a) and (b) are benchmarks presented for reference while case (d) corresponds to the well-known BLAST system in [4][16].

1) Interferer at the same data rate as desired signal: We examine the case of $T = 1$, $\beta = 1$, $\tau = 1/2$, and the nonsingular temporal interference correlation matrix shown in (24). Figs. 1 to 2 show the average SERs for training lengths $2N_t$ and $4N_t$, respectively. Comparing the LS

[2]For a nonsingular spatial interference correlation matrix, we set $N_r \leq L$.

detection (curve 7) with other methods, much lower symbol error rates can be achieved by using ordered MMSE detection as expected.

Comparing curves 5 and 6, we observe that for a training length of $4N_t$ symbols, gains can be obtained by estimating spatial interference correlation. However, shorter training lengths such as $2N_t$ produce inaccurate estimates of spatial interference correlation which in turn do not yield any benefit over assuming spatially white interference. As expected, we observe that the improvement by taking into account estimated spatial correlation increases with longer training lengths.

Examining curves 3 and 5 in Fig. 2, we observe that the improvement in taking temporal interference correlation into account in *channel estimation* is not significant. Moreover, this rate of improvement rapidly diminishes as the training length increases. This can be explained by noting that in estimating channel and spatial interference correlation matrices for temporally colored interference, the received signal vectors first undergo a transformation which temporally whitens the interference vectors as discussed in Section III-C. Since the temporal correlation in (24) drops quickly to zero after one time lag, the benefit in temporal whitening of interference vectors is not significant, especially for long training lengths.

By comparing curves 3 and 4 in Fig. 2, there is a slight improvement in using two-vector-symbol over one-vector-symbol detection. This implies that not much gain can be achieved by taking temporal interference correlation into account in *data detection* owing to the low temporal correlation. Due to better estimates of channel and interference spatial correlation matrices obtained with a longer training length, the performance gap between curves 3 and 4 should increase as the training length increases.

By comparing curves 4 and 6 in Fig. 2, we observe a 1.5dB gain in SIR obtained by estimating spatial interference correlation and taking explicit advantage of known temporal interference correlation in channel estimation and data detection using a training length of $4N_t$. About 1dB of that gain is due to the estimation of spatial interference correlation, and the remaining 0.5dB gain is due to exploiting temporal interference correlation in channel estimation and data detection.

2) Interferer at a lower data rate than desired signal: We examine the case of $T_I = 2T$, $T = 1$, $\beta = 1$, $\tau = 0.25$ and the temporal interference correlation matrix for training length $N = 6$ shown in (25). Recall that the temporal correlation matrix for the lower-data-rate-interferer case is singular. To avoid the singularity, the diagonal elements of Λ_N are increased by a small amount; hence, the temporal correlation matrix used for channel estimation may be modified to $\Lambda_N + \delta \mathbf{I}_N$ within the proposed framework. In our simulation, we chose $\delta = 0.01$.

The same set of average SER curves as in the same-data-rate-interferer case are simulated. Figs. 3 to 4 show the SERs for different training lengths. As in the case of the same-data-rate interferer, curve 7 illustrates the poor performance without ordered detection. Curves 5 and 6 suggest that for short training lengths it is better to estimate only the channel matrix and assume spatially white interference in data detection; however, for moderately long training lengths, gains can be obtained by estimating spatial interference correlation.

By examining curves 3 and 5 in Fig. 4, we observe that the improvement in taking temporal interference correlation into account in *channel estimation*, although larger than that in the same-data-rate-interferer case due to the high temporal correlation in the lower-data-rate-interferer case, is still not that significant.

In contrast to the same-data-rate-interferer case, curves 3 and 4 in Fig. 4 show that the improvement of two-vector-symbol over one-vector-symbol detection is significant due to the higher temporal interference correlation. This implies that significant gain can be achieved by taking known temporal interference correlation into account in *data detection* for the lower-data-rate-interferer case.

By comparing curves 4 and 6 in Fig. 4, for training length $4N_t$, there is a total of 4dB gain in SIR by estimating spatial interference correlation and taking advantage of known temporal interference correlation in channel estimation and data detection. About 3.5dB of the gain is due to exploiting temporal interference correlation in channel estimation and data detection.

3) Effect of model mismatch: With one interferer and a finite SNR, the interference-plus-noise statistics can only be approximately modelled using a Kronecker product. Here, we inves-

tigate when this approximation breaks down. We model thermal noise as a zero-mean circularly symmetric complex Gaussian vector with covariance matrix $\sigma^2 I_{N_r}$, i.e., independent from antenna to antenna, with noise power σ^2 on each antenna. We define interference-to-noise-power-ratio INR $= 10 \log P_I/\sigma^2$, where $P_I = 1$ in used in the simulations. For the case of an interferer at the same data rate and using a training length $4N_t$, we observe in curves 3 and 5 in Fig. 5 that at INRs below 17dB, taking interference temporal correlation into account appears not to be of benefit. Fig. 6 shows the corresponding comparison for the case of the lower-data-rate interferer. In this case, temporal correlation is larger and the decoupled model of interference-plus-noise statistics breaks down at INRs lower than 12dB.

4) Effect of exploiting spatial interference-plus-noise correlation: From the above results, temporal interference correlation, even if known, may not result in a performance benefit at lower INRs due to model mismatch. Therefore, we assess the benefit of taking only the *spatial* correlation of interference-plus-noise into account. As a reference, we compare performance to the case of assuming interference-plus-noise to be spatially white. With total interference power fixed, Fig. 7 compares the average SER for one (solid line) and two (broken line) interferers. In the case of two interferers, the interferers have equal power and random relative delays. Both desired and interfering users employ a $(5, 5)$ MIMO link, a total-interference-to-noise-ratio of 12dB, and the training length is $4N_t$. Both the desired and interfering users operate at the same data rate. Fig. 7 shows that for one interferer, there is 1.2dB gain over a wide range of SINRs by estimating the spatial correlation of interference-plus-noise. For the case of two equal-powered interferers, the corresponding gain in SINR is negligible.

VI. Conclusions

By modelling interference statistics as approximately temporally and spatially separable, we have investigated maximum-likelihood joint estimation of channel parameters and spatial interference correlation matrices. We have assessed the impact of temporal and spatial interference correlation on channel estimation and data detection. For training lengths of at least four times

the number of transmitting antennas, gains of around 1 dB are observed by estimating spatial interference correlation. We determine that an additional 0.5 to 3.0 dB in performance gain would result if known temporal correlation were exploited. For shorter training lengths, however, it is better to estimate only the channel matrix and assume spatially white interference in data detection. One source of temporal correlation occurs where a co-channel interferer operates at data rate lower than that of the desired user. Exploiting temporal interference correlation in *channel estimation* was found not to be of benefit. However, if temporal correlation is significant, as in case of lower-data-rate interference, significant performance gains by exploiting temporal interference correlation in *data detection* are theoretically possible. The minimum interference-to-noise (INR) levels where separable temporal and interference correlation statistics model was shown to break down and provide no benefit ranged from 12 or 17 dB, depending on the level of temporal correlation. Of more practical significance, it was shown that at a total INR of 12 dB, 1.2 dB of performance gain can be obtained over a wide range of SINRs by estimating spatial correlation only and neglecting temporal correlation.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 315–335, 1998.

[2] İ. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. Telecommun.*, pp. 585–595, Nov./Dec. 1999.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41–59, Autumn 1996.

[4] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, no. 1, pp. 14–15, Jan. 1999.

[5] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, pp. 744–765, Mar. 1998.

[6] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, pp. 1456–1467, July 1999.

[7] F. R. Farrokhi and G. J. Foschini, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Commun. Lett.*, pp. 85–87, Mar. 2001.

[8] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Simulation results for an interference-limited multiple-input multiple-output cellular system," *IEEE Commun. Lett.*, pp. 334–336, Nov. 2000.

[9] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Attainable throughput of an interference-limited multiple-input multiple-output (MIMO) cellular system," *IEEE Trans. Commun.*, pp. 1307–1311, Aug. 2001.

[10] Y. Song and S. D. Blostein, "MIMO channel capacity in co-channel interference," in *Proc. 21st Biennial Symposium on Commun.*, 2002, pp. 220–224.

[11] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inform. Theory*, pp. 3117–3128, Dec. 2002.

[12] R. S. Blum, "MIMO capacity with interference," *IEEE J. Select. Areas Commun.*, pp. 793–801, June 2003.

[13] M. Kang and M. Alouini, "Performance analysis of MIMO systems with co-channel interference over Rayleigh fading channels," in *Proc. IEEE Int. Conf. Communications*, 2002.

[14] B. Suard, A. F. Naguib, G. Xu, and A. Paulraj, "Performance of CDMA mobile communication systems using antenna arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1993.

[15] T. L. Marzetta, "BLAST training: estimating channel characteristics for high capacity space-time wireless," in *Proc. Allerton Conf. on Communication, Control, and Computing*, 1999, pp. 22–24.

[16] B. Hassibi, "A fast square-root implementation for BLAST," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, 2000.

[17] A. F. Naguib and A. Paulraj, "Performance of wireless CDMA with M-ary orthogonal modulation and cell site antenna arrays," *IEEE J. Select. Areas Commun.*, pp. 1770–1783, Dec. 1996.

[18] H. Lütkepohl, *Handbook of Matrices*. John Wiley & Sons, 1996.

[19] Y. Song, *Multiple-Input Multiple-Output Wireless Communication Systems with Cochannel Interference*, Ph.D. thesis, Dept. of Electrical and Computer Engineering, Queen's University, 2003.

[20] S. Bäro, G. Bauch, A. Pavilic, and A. Semmler, "Improving BLAST perfomance using space-time block codes and turbo decoding," in *Proc. IEEE Global Telecommun. Conf.*, 2000.

[21] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice-Hall, 1996.

[22] J. Proakis, *Digital Communications*. McGraw Hill, 1995.

[23] W. A. Gardner, *Introduction to Random Processes*. McGraw Hill, 1990.

[24] G. Long, F. Ling, and J. G. Proakis, "Fractionally-spaced equalizers based on singular values decomposition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1988.

~~[25] J. H. Winters, "Signal acquisition and tracking with adaptive arrays in the digital mobile radio system IS-54 with flat~~ fading," *IEEE Trans. Veh. Technol.*, pp. 377–384, Nov. 1993.

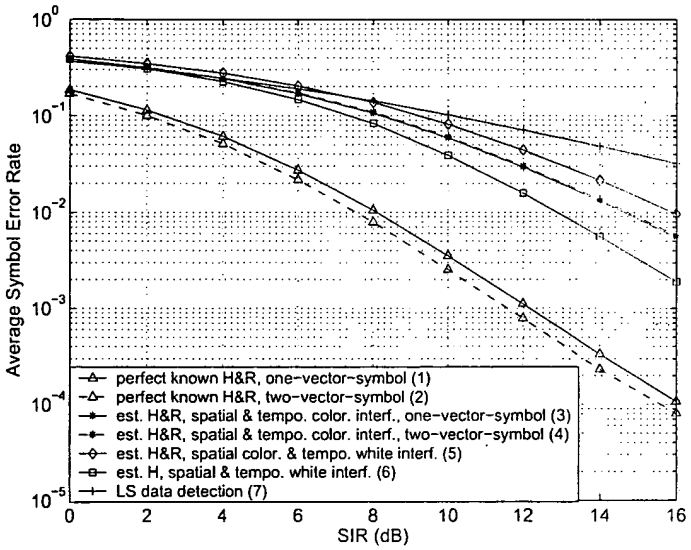


Fig. 1. Average symbol error rate vs. SIR with $N_t = N_r = 5$, $L = 6$, and training length $2N_t$ under independent Rayleigh fading. Both the desired and interfering users are at the same data rate.

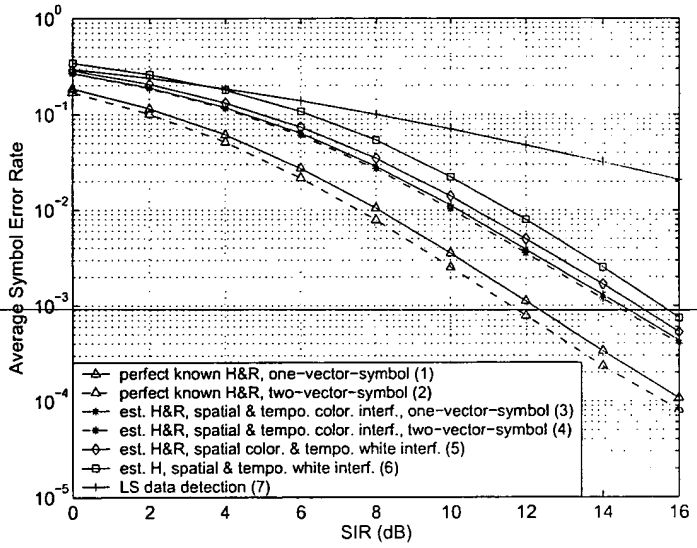


Fig. 2. Average symbol error rate vs. SIR with $N_t = N_r = 5$, $L = 6$, and training length $4N_t$ under independent Rayleigh fading. Both the desired and interfering users are at the same data rate.

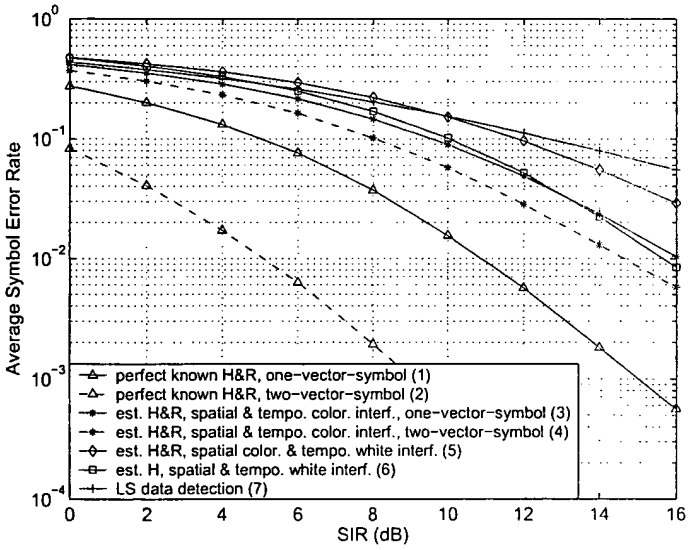


Fig. 3. Average symbol error rate vs. SIR with $N_t = N_r = 5$, $L = 6$, and training length $2N_t$ under independent Rayleigh fading. The data rate of the desired user is twice that of interfering user.

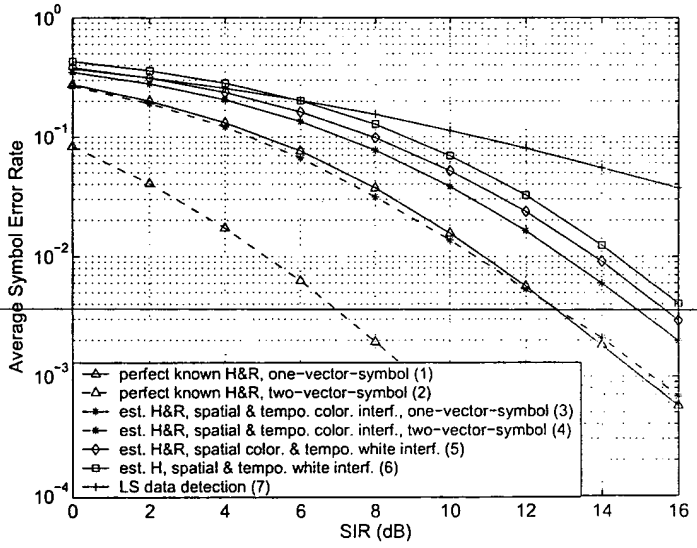


Fig. 4. Average symbol error rate vs. SIR with $N_t = N_r = 5$, $L = 6$, and training length $4N_t$ under independent Rayleigh fading. The data rate of the desired user is twice that of interfering user.

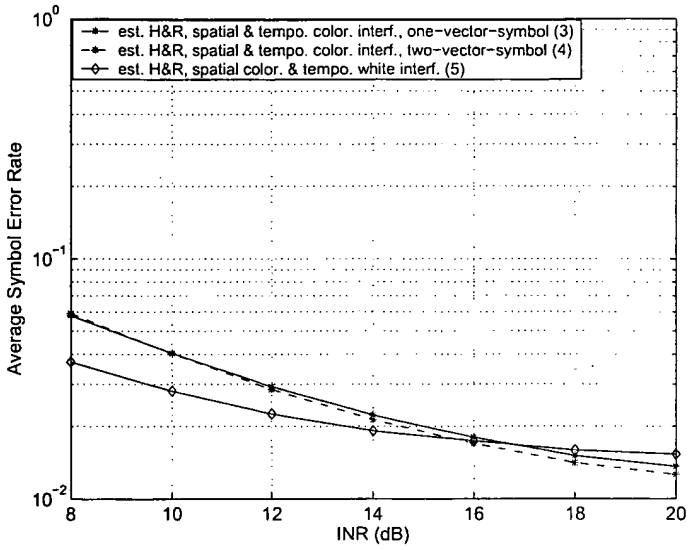


Fig. 5. Average symbol error rate vs. INR with $N_t = N_r = 5$, $L = 6$, SIR=10dB, and training length $4N_t$ under independent Rayleigh fading. Both the desired and interfering users are at the same data rate.

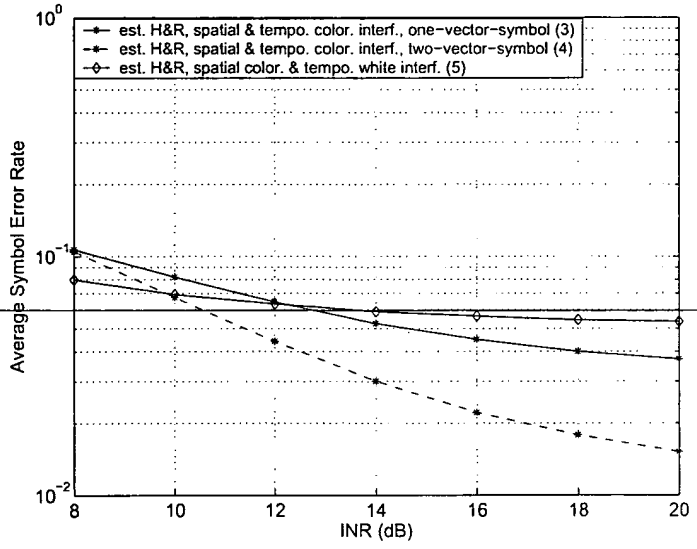


Fig. 6. Average symbol error rate vs. INR with $N_t = N_r = 5$, $L = 6$, SIR=10dB, and training length $4N_t$ under independent Rayleigh fading. The data rate of the desired user is twice that of interfering user.

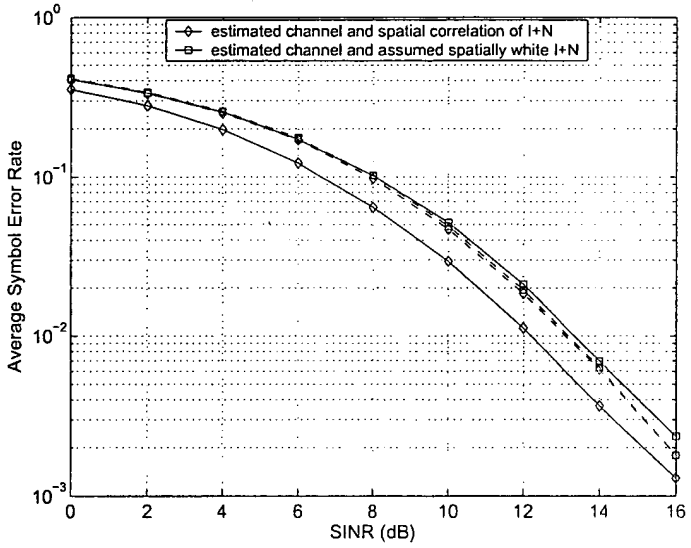


Fig. 7. The improvement of estimating spatial correlation of interference-plus-noise in practical systems. With total interference power fixed, the solid lines are for one interferer, and the broken lines are for two interferers. Both the desired and interfering users employ a $(5, 5)$ MIMO link, the same data rate, total-interference-to-noise-ratio of 12dB, and training length of $4N_t$.

Unique Features of Subspace Processing for Adaptive Radar in Inhomogeneous Environments[1]

William L. Melvin
Georgia Tech Research Institute
Sensors & Electromagnetic Applications Laboratory
Atlanta, GA 30332-0857 USA

Abstract – **Space-time adaptive processing (STAP) is a key technology for improved detection of airbreathing and ground moving targets from airborne and spaceborne radar platforms. Subspace methods are central to STAP algorithm development and provide important insight into adaptive radar signal processing. In this paper, we consider unique aspects of subspace methods to address sample support requirements and training approaches for covariance matrix estimation. Furthermore, we discuss subspace characterizations for inhomogeneous signal environments and validate the proposed models with measured airborne radar data. It is shown herein that inhomogeneity degrades STAP performance and that subspace methods intrinsically serve as a cornerstone for robust STAP algorithm development.**

I. Introduction

Space-time adaptive processing (STAP) is a multi-dimensional, adaptive filtering approach with application to airborne [1] and spaceborne radar, as well as communication and sonar systems. In the case of airborne radar, STAP exploits diversity among measurements in azimuth, elevation, slow-time (Doppler) and fast-time (range) to suppress correlated clutter and jamming signals while enhancing target signal components. Thus, STAP enables the detection of weak targets. The terminology adaptive implies that critical statistical quantities are unknown, but are estimated from available secondary data. When the secondary data appear inhomogeneous, erroneous covariance matrix estimates lead to STAP performance degradation with respect to the optimal condition. The deleterious effect of inhomogeneity on STAP has previously been reported based on the analysis of measured radar data from the US Air Force sponsored Multichannel Airborne Radar Measurements (MCARM) effort and DARPA's Mountaintop program [2-3].

While clutter inhomogeneity is a principal concern for adaptive airborne surveillance radar, other issues abound in the research literature, including the development of novel, robust algorithms. Recently reported subspace-based STAP methods hold particular interest since such approaches may provide results closer to optimal than competing techniques [4-6]. Subspace concepts also suggest STAP heuristics with lower computational burden [7]. Yet, beyond algorithm development, subspace methods possess unique attributes essential for advancing adaptive airborne radar. In this paper, we use subspace principles to discuss localized training, characterize inhomogeneous signal environments and develop techniques to enhance STAP capability.

II. Multichannel Airborne Radar Signal Processing

A. STAP Basics

A seminal paper by Brennan and Reed discusses the basic theory of STAP for airborne radar [1]. We provide a brief overview. Consider a pulse Doppler radar with M spatial channels radiating N pulses. Next, define the vector of all complex baseband observations within the coherent processing interval (CPI), corresponding to range cell k, as

$$x_k = \left[x_{s,k}(1) \;\; x_{s,k}(2) \;\; \cdots \;\; x_{s,k}(N-1) \;\; x_{s,k}(N) \right]^T . \quad \textbf{(1)}$$

$x_{s,k}(n) \in C^{Mx1}$ is the *spatial snapshot* comprised of all spatial samples for the nth pulse, whereas we refer to $x_k \in C^{MNx1}$ as the *space-time data vector*. In STAP, we compute the adaptive filter output as

$$y_k = \hat{w}_k^H x_k \; ; \quad \hat{w}_k = \alpha \, \hat{R}_k^{-1} s_{s-t}(\theta, f_d) , \quad \textbf{(2)}$$

where $\hat{w}_k \in C^{MNx1}$ is the adaptive weight vector, $\hat{R}_k \in C^{MNx1}$ is the sample covariance matrix approximating the true, unknown interference covariance matrix, $R_k = E[x_{k/H_0} x_{k/H_0}^H]$, $s_{s-t}(\theta, f_d)$ approximates the target steering vector for angle θ and Doppler f_d, and α is a constant. The choice for $\hat{w}_k$ in (2) attempts to maximize signal-to-interference plus noise ratio (SINR). Generally, one employs the maximum likelihood estimate (MLE)

$$\hat{R}_k = \frac{1}{L} \sum_{m=1}^{L} x_m x_m^H , \quad \textbf{(3)}$$

[1] Presented at the 31st Southeastern Symposium on System Theory (SSST99), 21-23 March 1999, Auburn University, Alabama.

to compute the sample covariance matrix. The Reed-Mallett-Brennan (RMB) rule governs the use of (3).

Reed-Mallett-Brennan (RMB) Rule [8]: Assume the secondary data, $\{x_m\}_{m=1}^{L}$, are independent and identically distributed (iid). Then, equation (3) minimally requires $L \approx 2 \cdot dim(x_k)$ if the expected SINR of the adaptive processor output is to be within 3 dB of optimal.

A limited number of potentially inhomogeneous secondary data exacerbates use of (3).

B. Basic Theory of Subspace-Based STAP Techniques

Subspace methods are appealing because they allow us to consider multidimensional signal vectors in terms of a limited size, well-defined basis. The eigenvectors of the covariance matrix serve as this basis. Observe that $\hat{R}_k$ is Hermitian and positive definite. For this reason, a unitary matrix Q_k exists such that

$$\hat{R}_k = Q_k \Lambda_k Q_k^H = \sum_{m=1}^{MN} \lambda_k(m)\, q_k(m)\, q_k^H(m)$$
$$\Lambda_k = diag\left(\lambda_k(m)\right)_{m=1}^{MN}; \qquad (4)$$
$$Q_k = \left[q_k(1)\ q_k(2)\ \cdots\ q_k(MN)\right],$$

where $\lambda_k(m)$ is an eigenvalue of $\hat{R}_k$ with related eigenvector $q_k(m)$. Let $\lambda_k(1) \geq \lambda_k(2) \geq \cdots \geq \lambda_k(MN)$, and note that $\{q_k(m)\}_{m=1}^{MN}$ defines an orthonormal set. The larger values of $\lambda_k(m)$ define the *interference subspace*, whereas the smaller values define the *noise subspace*. The interference-plus-noise covariance matrix is generally of low numerical rank [4-6, 7].

We may express the optimal weight vector as [4]

$$w_k^H = \frac{1}{\lambda_0}\left(s_{s\text{-}t}^H - \sum_{m=1}^{MN} \frac{\lambda_k(m) - \lambda_0}{\lambda_k(m)}\, \mu_k(m)\, q_k^H(m) \right), \qquad (5)$$

where λ_0 represents the noise floor eigenvalue, $\mu_k(m) = s_{s\text{-}t}^H q_k(m)$ is the projection of the mth eigenvector onto the quiescent pattern defined by $s_{s\text{-}t}^H$ and $Q_k Q_k^H = I_{MN}$. From (5) we observe that the weight vector equals the quiescent response minus the weighted sum of eigenvectors. Ideally, noise eigenvalues result in no subtraction, whereas interference eigenvectors lead to a notch in the quiescent pattern. Often, the noise subspace appears perturbed due to the limited sample support. This perturbation results in raised sidelobe levels due to the addition of random terms in (5). Constructing the weight vector from principal eigenvectors, or applying diagonal loading to the sample covariance matrix, i.e., adding an identity matrix times a constant to $\hat{R}_k$, improves the sidelobe response at the expense of reduced null depth.

The recent literature appears to categorize subspace-based STAP algorithms as follows. Principal component approaches choose the eigenbasis corresponding to the largest eigenvalues to construct the weight vector [4]. In contrast, metric-based approaches select the eigenbasis to optimize a given cost function [6]. This eigenbasis may span both noise and interference subspaces. Principal component inverse methods construct a weight vector spanning the noise subspace [5]. Since the noise and interference subspaces appear orthogonal, constructing the weight vector as a linear combination of noise eigenvectors anhilates correlated clutter lying in the orthogonal interference subspace.

III. Localized Training

As already mentioned, airborne radar signal environments can appear quite inhomogeneous. Conditions leading to instances of clutter inhomogeneity include shadowing, spatially-varying clutter reflectivity and interfaces between dominant clutter types (e.g., land-sea interface), clutter discretes, moving scatterers, and weather phenomena. Additionally, chaff and coherent repeater jamming are cases of induced inhomogeneity. In the case of spatially-varying clutter and interfaces or edges, it is desirable to locally train the adaptive processor to best capture interference statistics nearest the primary test data. Even if all data were *iid*, the product of M and N can appear sufficiently large such that use of (3) does not yield acceptable performance. We may exploit the low numerical rank of the clutter covariance matrix to reduce the required sample support.

Dominant Subspace Estimation: The sample support required to estimate the dominant components of the interference covariance matrix is on the order of $2 \cdot rank(R_{k,I})$, where $R_{k,I}$ is the interference covariance matrix. Exact dominant subspace information enables optimal performance as seen via (5).

Two issues to consider are as follows. Computing the subspace information can be numerically burdensome. A heuristic proposed in [7] ameliorates this situation. Similarly, beamspace and Doppler transformations often serve as suitable rank-reducing approaches. Diagonal loading can be used to improve poor adaptive receive sidelobe response, but this is a trade-off since it impacts attainable null depth and the appropriate level of loading is heuristic [10]. In Figure 1, we compare the eigenspectra of the interference-plus-noise covariance matrix for the case of a linear array and varying *iid* sample support. Observe that principal eigenvalues are well represented, independent of sample support, whereas estimating the noise eigenvalues is a much more difficult task. Figure 2 shows the adapted patterns for several cases of varying sample support. Notice that the poor noise eigenvalue estimates lead to perturbed sidelobe levels, as indicated by (5). Choosing only principal components to construct the weight vector removes raised sidelobe levels. Diagonal loading offers a similar effect by compressing the noise eigenvalues and improving the covariance matrix condition number.

In summary, subspace processing requires lower sample support, which in turn allows localized processing. This is a necessary step towards the best possible performance when operating in inhomogeneous clutter scenarios.

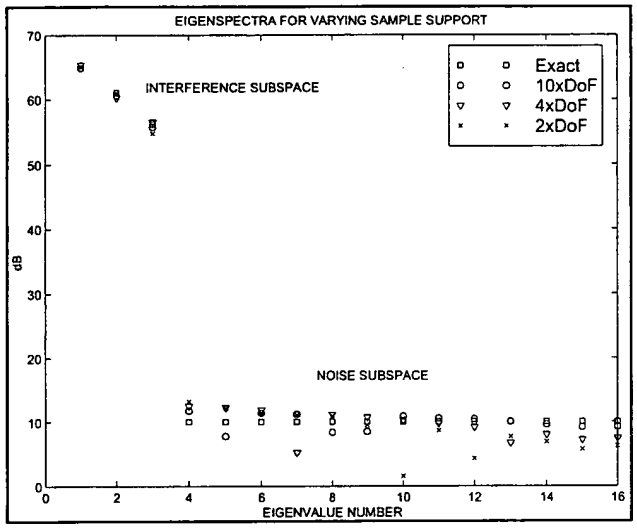


Fig. 1. Eigenspectra for varying sample support measured in terms of degrees of freedom (DoF). Three noise jammers are present.

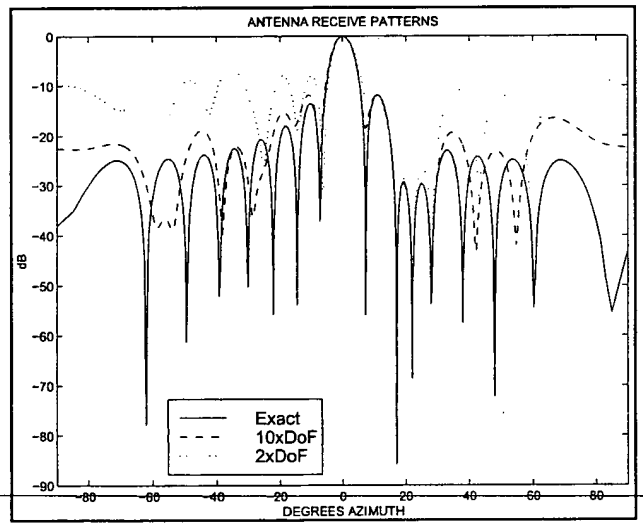


Fig. 2. Adapted receive patterns, corresponding to the example shown in Figure 1, for varying sample support.

IV. General Multichannel Model For Clutter Inhomogeneity

Using subspace principles, we review a simple model for clutter inhomogeneity. This model is discussed in more detail in [2]. The low rank clutter-plus-noise covariance matrix, with σ_n^2 representing the noise variance, takes the form

$$R_k = \sum_{\ell=1}^{P} \tilde{\sigma}_k^2(\ell)\, \tilde{s}_k(\ell)\, \tilde{s}_k^H(\ell) + \sigma_n^2\, I_{MN}\,, \qquad (6)$$

under simplifying assumptions. In this case, $\tilde{\sigma}_k^2(\ell)$ and $\tilde{s}_k(\ell)$ are unknown variables representing the power and space-time steering vector for the ℓ^{th} of P principal interference sources. It is the variation of $\tilde{\sigma}_k^2(\ell)$ and $\tilde{s}_k(\ell)$ from realization to realization which defines a condition of inhomogeneity. The effect is best understood by examining the impact on covariance structure. Considering principal eigenvector terms, we see from the eigenanalysis equation,

$$\hat{R}_k\, q_k(m) = \lambda_k(m)\, q_k(m) \qquad (7)$$

that

$$q_k(m) = \sum_{\ell=1}^{P} \left(\frac{\tilde{\sigma}_k^2(\ell)\, \eta_k(\ell,m)}{\lambda_k(m) - \sigma_n^2} \right) \tilde{s}_k(\ell)\,; \qquad (8)$$

$$\eta_k(\ell,m) = \tilde{s}_k^H(\ell)\, q_k(m)\,,$$

for the principal eigenvectors. $\eta_k(\ell,m)$ represents the projection of the m^{th} eigenvector onto the ℓ^{th} interference steering vector.

Equation (8) is very interesting, showing that each principal eigenvector spans all interference steering vectors. Thus, inhomogeneity necessarily affects all eigenvalues and eigenvectors to a certain degree. This indicates a change in covariance structure. The preceding description of inhomogeneity is useful since we can apply it to measured data using the singular value decomposition (SVD).

V. Measured Data: Examples of Inhomogeneity

We now show some examples using measured data from the MCARM program to validate our view of inhomogeneity. The measured data is taken from MCARM flight 5, acquisition 575. This data is publicly available and may be found at [11], along with more detail on the MCARM system. Figure 3 shows the eigenbeams of the spatial covariance matrix in the Doppler filter corresponding to a radial velocity of 42 miles per hour for 136 samples taken from range cell 245 to range cell 285. Eigenbeams are essentially Fourier transforms of the covariance matrix eigenvectors. As predicted by (8), they commonly point towards an interference source. Figure 4 shows the eigenbeams for the covariance matrix computed from the subset of secondary data, within the original set of 136 range samples, determined to be most homogeneous. The test statistic used to determine homogeneity is mentioned briefly in the following section and in [2, 9]. Observe that the eigenbeam closest to zero degrees is absent. We attribute this eigenbeam to moving scatterers. In terms of (8), moving scatterers in the secondary

data set result in an added principle component with eigenvector $q_k(0) \approx \tilde{s}_m(0)/\|\tilde{s}_m(0)\|_2^{1/2}$, where $\tilde{s}_m(0)$ is the added steering vector due to the moving scatterer in the secondary data, and an eigenvalue determined by the reflected power of the inhomogeneity de-weighted by the number of range samples in the MLE. Note that adding samples in the MLE does not necessarily improve performance because other inhomogeneity may enter the secondary data set.

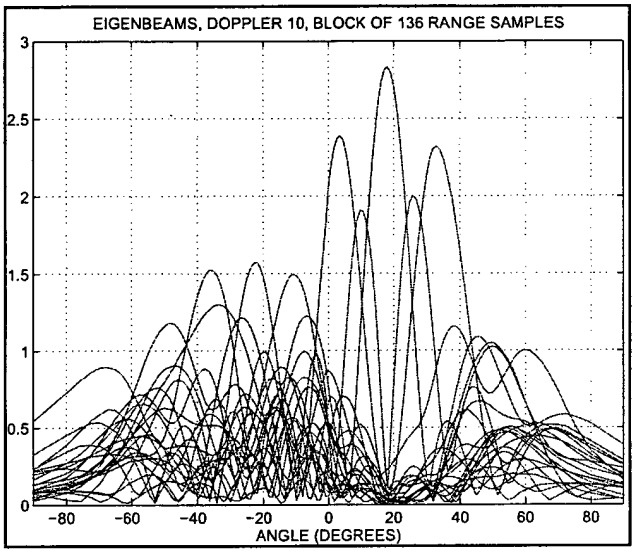


Fig. 3. MCARM data. Eigenbeams for block of consecutively selected secondary data within Doppler filter 10.

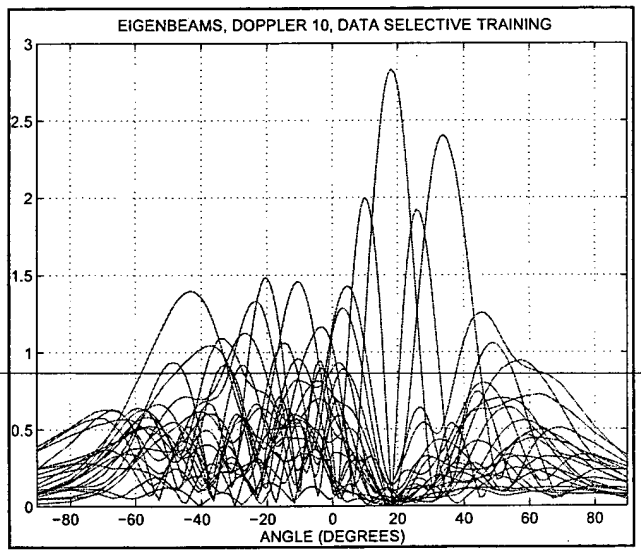


Fig. 4. MCARM data. Eigenbeams for selectively determined secondary data chosen as a subset of the same data swath used to produce Figure 3.

Figure 5 shows the eigenspectra for the MCARM data over Doppler filters 9 through 16 (128 pulse FFT) with varying sample support. Secondary data are chosen from a near-in range, with the window expanding to accommodate greater sample support. Despite the use of sensitivity time control (STC), shadowing and edge effects lead to underestimated eigenvalue magnitude. In terms of (6-8), the observed shadowing leads to underestimated eigenvalues $\hat{\lambda}_k(m) \approx \epsilon\, \lambda_k(m)$ with $\epsilon < 1$, and consequently resulting in a condition of undernulled clutter.

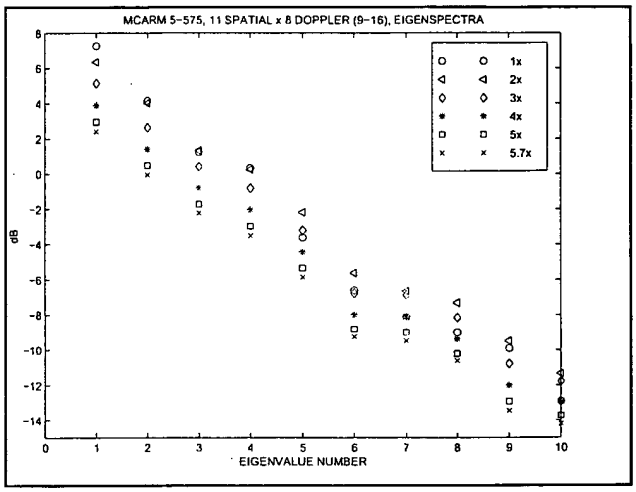


Fig. 5. MCARM Data. Eigenspectra across Doppler filters 9 through 16 showing impact of shadowing. The legend inputs, "mx", indicate sample support in terms of available degrees of freedom.

VI. Approaches For Robust Performance

Assumptions intrinsic to STAP development, namely the *iid* assumption of the secondary data set, can be routinely violated in practical scenarios due to inhomogeneity. Consider the simple test statistic of the form

$$\gamma_m = x_m^H R_T^{-1} x_m \,. \qquad (9)$$

Basically, (9) determines how well $R_T^{-1/2}$ whitens the input data vector and is a chi-squared random variable. We may consider data vectors similarly whitened, thus having similar values of γ_m, to be homogeneous. In terms of eigenstructure, the expected value of (9) measures the sum of the ratio of the data vector's unknown eigenvalues to the eigenvalues of R_T [2]. Taking the expected value of (9), after substituting the spectral decomposition for R_T and the Karhunen-Loeve eigenbasis representation for x_m, validates the preceding statement. Applying (9) to MCARM data, we find that the choice of secondary data strongly influences attainable detection performance. This is shown in Figure 6, where the top plot shows the output of the adaptive processor (adaptive nulling in Doppler filter 10, sometimes called factored time-space (FTS)) with continuous weight updates using an MLE computed from range cells taken via a symmetric window (SW) about the primary data cell. The number of samples selected is twice the processor's degrees of freedom. In contrast, the bottom plot depicts performance when computing an MLE from an equal number of the most homogeneous samples in this same region. Equation (9) assesses relative homogeneity of the secondary

data in this case. In this instance, an injected target at range cell 290 with - 19 dB SNR is clearly visible when the processor carefully selects training data.

The example in Figure 6 shows that clutter inhomogeneity can seriously degrade STAP capability. Other approaches for robust STAP performance should be developed to overcome some of the effects of clutter inhomogeneity. These approaches may not necessarily involve training sample selection. Potential approaches include pre-filtering and modification of secondary data, use of knowledge bases and application of terrain mapping data. It is likely that the best approach depends on the form of clutter inhomogeneity. This is a topic for further research.

VII. Summary

In this paper we describe the utility of subspace processing in light of inhomogeneous clutter. Specifically, through subspace techniques, we find that localized training is possible with sample support on the order of twice the clutter rank. Furthermore, subspace methods provide a framework for characterizing clutter inhomogeneity and the resulting changes one might expect in covariance matrix structure. Finally, we mention an approach for quantifying covariance structure via a test statistic that is best understood in terms of eigenanalysis. Examples using measured data from the MCARM program show that inhomogeneous clutter leads to significantly degraded STAP performance. Furthermore, the MCARM data corroborates proposed subspace characterizations of clutter inhomogeneity.

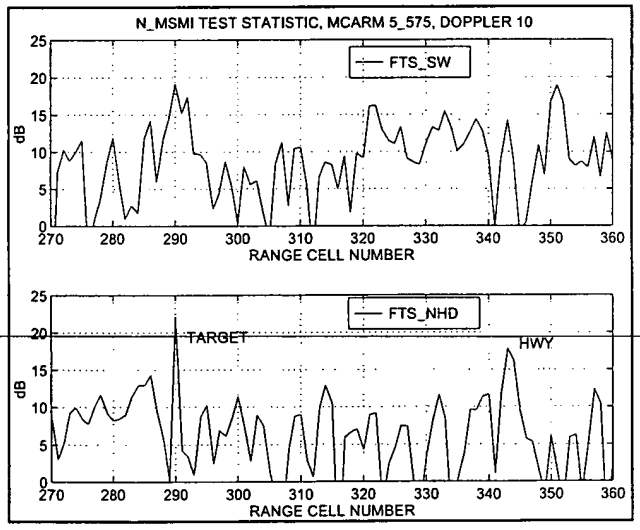


Fig. 6. Example of the influence of sample selection on detection performance. "FTS_SW" indicates adaptive nulling in Doppler 10, using a symmetric window (SW) to choose secondary data. "FTS_NH" uses data dependent secondary data selection to yield improved performance of the adaptive filter. The notation "N_msmi" indicates the adaptive filter output is normalized by $1/s_{s,t}^{H} R_k^{-1} s_{s,t}$.

VIII. Acknowledgments

This work was supported under FY 1999 Georgia Tech Research Institute IR&D funding.

IX. References

[1] L.E. Brennan and I.S. Reed, "Theory of adaptive radar", *IEEE Trans. AES*, Vol. 9, No. 2, March 1973, pp. 237-252.

[2] W.L. Melvin , "Eigenbased Modeling of Nonhomogeneous Airborne Radar Environments," in *Proc. of the 1998 IEEE Radar Conference*, Dallas, TX, May 1998, pp. 171-176.

[3] D.J. Rabideau and A.O. Steinhardt, "Improving the Performance of Adaptive Arrays in Nonstationary Environments Through Data-Adaptive Training," in *Proc. 30th Asilomar Conf.*, Pacific Grove, CA, 3-6 November 1996, pp. 75-79.

[4] W.F. Gabriel, "Using spectral estimation techniques in adaptive processing antenna systems," *IEEE Trans. AP*, Vol. 34, No. 3, March 1986, pp. 291-300.

[5] A. Haimovich, "The eigencanceler: adaptive radar by eigenanalysis methods," *IEEE Trans. AES*, Vol. 32, No. 2, April 1996, pp. 532-542.

[6] J.S. Goldstein and I.S. Reed, "Theory of partially adaptive radar," *IEEE Trans. AES*, Vol. 33, No. 4, Oct. 1997, pp. 1309-1325.

[7] A.W. Bojanczyk and W.L. Melvin, "Simplifying the Computational Aspects of STAP," in *Proc. of the 1998 IEEE National Radar Conference*, Dallas, TX, May 1998, pp. 123-128.

[8] I.S. Reed, J.D. Mallett, and L.E. Brennan, "Rapid convergence rate in adaptive arrays", *IEEE Trans. AES*, Vol. 10, No. 6, November 1974, pp. 853-863.

[9] P. Chen, W. Melvin and M. Wicks, "Screening among multivariate normal data," to appear in *Journal of Multivariate Analysis*, January 1999.

[10] B.D. Carlson, "Covariance Matrix Estimation Errors and Diagonal Loading in Adaptive Arrays," *IEEE Trans. AES*, Vol. 24, No. 4, July 1988, pp. 397-401.

[11] Pertinent MCARM information and data is available via the internet. Go to http://sunrise.deepthought.rl.af.mil/. Questions regarding this web site should be addressed to the designated US Air Force Research Laboratory (Rome Site) point of contact.

Interference Mitigation in STAP Using the Two-Dimensional Wold Decomposition Model

Joseph M. Francos, *Senior Member, IEEE*, and Arye Nehorai, *Fellow, IEEE*

***Abstract*—We propose a novel parametric approach for modeling, estimation, and detection in space-time adaptive processing (STAP) radar systems. The proposed parametric interference mitigation procedures can be applied even when information in only a single range gate is available, thus achieving high performance gain when the data in the different range gates cannot be assumed stationary. The model is based on the Wold-like decomposition of two-dimensioanl (2-D) random fields. It is first shown that the same parametric model that results from the 2-D Wold-like orthogonal decomposition naturally arises as the physical model in the problem of space-time processing of airborne radar data. We exploit this correspondence to derive computationally efficient fully adaptive and partially adaptive detection algorithms. Having estimated the models of the noise and interference components of the field, the estimated parameters are substituted into the parametric expression of the interference-plus-noise covariance matrix. Hence, an estimate of the fully adaptive weight vector is obtained, and a corresponding test is derived. Moreover, we prove that it is sufficient to estimate only the spectral support parameters of each interference component in order to obtain a projection matrix onto the subspace orthogonal to the interference subspace. The resulting partially adaptive detector is simple to implement, as only a very small number of unknown parameters need to be estimated, rather than the field covariance matrix. The performance of the proposed methods is illustrated using numerical examples.**

***Index Terms*—Airborne radar, clutter, detection, evanescent fields, interference mitigation, jamming, STAP, two-dimensional random fields, Wold decomposition.**

I. Introduction

WE PROPOSE a new approach for parametric modeling and estimation of space-time airborne radar data, based on the two-dimensional (2-D) Wold-like decomposition of random fields. Most interestingly, the proposed parametric estimation algorithms of the interference components provide new tools to estimate and mitigate the Doppler ambiguous clutter. The algorithms we develop enable estimation of the interference signals using the observations in *only* a single range gate. This property makes the proposed method particularly suitable for nonstationary clutter and jamming environments. Our modeling approach also provides a new analytical insight into the space-time adaptive processing (STAP) problem.

The goal of STAP is to manipulate the available data to achieve high gain at the target's angle and Doppler and maximal mitigation along both the jamming and clutter lines. Because the interference covariance matrix is unknown *a priori*, it is typically estimated using sample covariances obtained from averaging over a few range gates. Next, a weight vector is computed from the inverse of the sample covariance matrix [1]–[5]. It is shown in [6] that the dominant eigenvectors of the space-time covariance matrix contain all the information required to mitigate the interference. Thus, the weight vector is constrained to be in the subspace orthogonal to the dominant eigenvectors. In [8], a reduced-rank constant false alarm (CFAR) detection test is developed, assuming the dominant eigenvectors of the interference are known, and in [9], a multistage partially adaptive CFAR detection algorithm is introduced. In [17], an approach that bypasses the need to estimate the covariance matrix is presented: The data collected in a single range gate is employed to obtain a least-squares estimate of the signal power at each hypothesized direction of arrival, through evaluation of a weight vector constrained to null the unknown interference and noise. In [18], a simple *ad hoc* model of the clutter signal and covariance matrix is proposed. The model represents the spectral density of the clutter as a sum of Gaussian-shaped humps along the support of the clutter ridge. In [19], this model is employed to estimate the clutter covariance matrix from the data observed in a single range gate.

In this paper, we adopt the 2-D Wold-like decomposition of random fields [10] as the parametric model of the observed data. Employing this model, we derive computationally efficient algorithms useful for parametrically estimating both the jamming and clutter fields. The estimation procedure we propose is capable of estimating the interference parameters from the information in a single range gate. Hence, no averaging over a few range gates is required. This property provides significant advantage in the practical case where data in the different range gates is nonstationary. Having estimated the interference terms parametric models, their covariance matrix can be evaluated based on the estimated parameters. Moreover, the problem of evaluating the rank of the low-rank covariance matrix of the interference is solved as a byproduct of obtaining the parametric estimates of the interference components. Once the parametric models of the interference components have been estimated, several alternative detection procedures

Manuscript received April 7, 2002; revised February 27, 2003. This work was supported by the Air Force Office of Scientific Research under Grants F49620-00-1-0083 and F49620-02-1-0339, the National Science Foundation under Grant CCR-0105334, and the Office of Naval Research under Grant N00014-01-1-0681. The associate editor coordinating the review of this paper and approving it for publication was Prof. Zhi Ding.

J. M. Francos is with the Electrical and Computer Engineering Department, Ben-Gurion University, Beer-Sheva 84105 Israel (e-mail: francos@ee.bgu.ac.il).

A. Nehorai is with the Electrical and Computer Engineering Department, University of Illinois at Chicago, Chicago, IL 60607-7053 USA (e-mail: nehorai@ece.uic.edu).

Digital Object Identifier 10.1109/TSP.2003.816883

are available. In this paper, we present two such methods: the parametric fully-adaptive processing and the parametric partially-adaptive processing.

The paper is organized as follows. In Section II, we briefly summarize the main results of the 2-D Wold-like decomposition and the resulting random field model. Next, in Section III, the correspondence between this model and the physical model of the STAP data is identified. In Section IV, we elaborate on the parametric representation of the covariances of the different components of the random field. The estimation algorithm of the random field parametric model is presented and analyzed in Section V. After the method for estimating the parametric models of the different components of the data field has been established, we present the parametric fully adaptive processing method and the computationally more efficient parametric partially adaptive processing method in Sections VI and VII, respectively. The performance of both methods is illustrated using synthetic data examples. We summarize our conclusions in Section VIII.

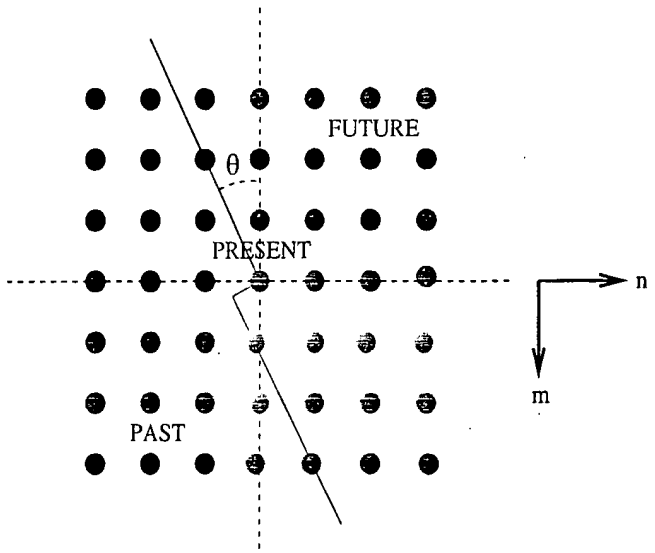


Fig. 1. Rational nonsymmetrical half-plane support; example with $a = 2$ and $b = -1$.

II. Random Field Model

In this section, we briefly review the 2-D Wold-like decomposition of random fields and the resulting random field model. In the next section, the applicability of this model to STAP data will be explained. It is shown in [10] that any 2-D regular and homogeneous discrete random field can be represented as a sum of two mutually orthogonal components: a *purely indeterministic* (unpredictable in the mean-square sense) field and a *deterministic* (predictable in the mean-square sense) one. The purely indeterministic component has a unique white innovations driven nonsymmetrical half-plane (NSHP) moving average representation. The deterministic component is further orthogonally decomposed into a *harmonic* field and a countable number of mutually orthogonal *evanescent* fields. This decomposition results in a corresponding decomposition of the spectral measure of the regular random field into a countable sum of mutually singular spectral measures. The purely indeterministic component has an absolutely continuous spectral distribution function. The spectral measure of the deterministic component is singular with respect to the Lebesgue measure, and therefore, it is concentrated on a set of Lebesgue measure zero in the frequency plane. It is shown in [12] that under some mild assumptions (that always hold in practice), the spectral supports of the different evanescent components have the form of lines whose slope is a rational number.

More specifically, let $\{y(n, m), (n, m) \in \mathcal{Z}^2\}$ be a complex valued, regular, homogeneous random field. Then, $y(n, m)$ can be uniquely represented by the orthogonal decomposition

$$y(n, m) = w(n, m) + v(n, m). \tag{1}$$

The field $\{v(n, m)\}$ is a deterministic random field. The field $\{w(n, m)\}$ is purely indeterministic and has a unique white innovations driven moving average representation, which is given by

$$w(n, m) = \sum_{(0,0) \preceq (k,\ell)} b(k, \ell) u(n - k, m - \ell) \tag{2}$$

where $\sum_{(0,0) \preceq (k,\ell)} b^2(k, \ell) < \infty$; $b(0, 0) = 1$, and $\{u(n, m)\}$ is the innovations field of $\{y(n, m)\}$. The notation $\preceq$ implies that the weighted summation includes $u(n, m)$ and all the samples in its "past," where the past is defined with respect to any selected NSHP total-ordering on the 2-D lattice (see, for example, Fig. 1).

We call a 2-D deterministic random field $[e_o(n, m)]$ *evanescent w.r.t. the NSHP total-order* o if it spans a Hilbert space identical to the one spanned by its *column-to-column innovations* at each coordinate (n, m) (w.r.t. the total-order o). The deterministic field column-to-column innovation at each coordinate $(n, m) \in \mathcal{Z}^2$ is defined as the difference between the actual value of the field and its projection on the Hilbert space spanned by the deterministic field samples in all previous columns.

It is possible to define [10] a family of NSHP total-order definitions such that the boundary line of the NSHP has a rational slope. A NSHP of this type is called *rational nonsymmetrical half-plane* (RNSHP), (see, for example, Fig. 1). Let a and b be two coprime integers, such that both $a, b \neq 0$. The slope of the RNSHP is then given by $-a/b$ (and $\cot\theta = -b/a$). For the case where $a = 0$, the RNSHP is uniquely defined by setting $b = 1$. (For the case where $b = 0$, the RNSHP is uniquely defined by setting $a = 1$.) We denote by O the set of all possible RNSHP definitions on the 2-D lattice (i.e., the set of all NSHP definitions in which the boundary line of the NSHP has a rational slope). The introduction of the family of RNSHP total-ordering definitions results in the following countably infinite orthogonal decomposition of the deterministic component of the random field:

$$v(n, m) = p(n, m) + \sum_{(a,b) \in O} e_{(a,b)}(n, m). \tag{3}$$

The random field $\{p(n, m)\}$ is *half-plane deterministic*, i.e., it has no column-to-column innovations w.r.t. any RNSHP total-ordering definition. The field $\{e_{(a,b)}(n, m)\}$ is the evanescent component that generates the column-to-column innovations of

the deterministic field w.r.t. the RNSHP total-ordering definition $(a, b) \in O$.

Hence, if $\{y(n, m)\}$ is a 2-D regular and homogeneous random field, then $y(n, m)$ can be uniquely represented by the orthogonal decomposition

$$y(n, m) = w(n, m) + p(n, m) + \sum_{(a,b)\in O} e_{(a,b)}(n, m). \quad (4)$$

In the following, all spectral measures are defined on the square $K = [-1/2, 1/2] \times [-1/2, 1/2]$. It is shown in [10] and [11] that the spectral measures of the decomposition components in (4) are mutually singular. The spectral distribution function of the purely indeterministic component is absolutely continuous, whereas the spectral measures of the half-plane deterministic component and of all the evanescent components are concentrated on a set of Lebesgue measure zero in K. A model for the evanescent field that corresponds to the RNSHP defined by $(a, b) \in O$ is given by

$$\begin{aligned} e_{(a,b)}(n, m) &= \sum_{i=1}^{I^{(a,b)}} e_i^{(a,b)}(n, m) \\ &= \sum_{i=1}^{I^{(a,b)}} s_i^{(a,b)}(na + mb) \\ &\quad \cdot \exp\left(j2\pi\nu_i^{(a,b)}(nc + md)\right) \end{aligned} \quad (5)$$

where c and d are coprime integers satisfying $ad - bc = 1$. For the case where $(a, b) = (0, 1)$, we have $(c, d) = (1, 0)$, and for $(a, b) = (1, 0)$, we have $(c, d) = (0, 1)$. The 1-D purely indeterministic, complex-valued processes $\{s_i^{(a,b)}(na + mb)\}$ and $\{s_j^{(a,b)}(na + mb)\}$ are zero-mean and mutually orthogonal for all $i \neq j$. Hence, the "spectral density function" of each evanescent field has the form of a sum of 1-D delta functions that are supported on lines of rational slope in the 2-D spectral domain. The amplitude of each of these delta functions is determined by the spectral density of the 1-D modulating process. Since the spectral density of the modulating process can rapidly decay to zero, so will the "spectral density" of the evanescent field, hence, the name "evanescent." Since interchanging the roles of past and future in any total-order definition amounts to substituting $\nu_i^{(a,b)}$ by $-\nu_i^{(a,b)}$ in the model (5), we assume without limiting the generality of the derivation that $a \geq 0$, and b can assume any integer value.

One of the half-plane-deterministic field components, which is of prime importance in the STAP problem is the harmonic random field

$$h(n, m) = \sum_{p=1}^{P} C_p \exp\left(j2\pi(n\omega_p + m\nu_p)\right) \quad (6)$$

where the C_ps are mutually orthogonal random variables, and (ω_p, ν_p) are the spatial frequencies of the pth harmonic.

III. STAP Model and the 2-D Wold Decomposition

The random field parametric model that results from the 2-D Wold-like orthogonal decomposition naturally arises as the physical model in the problem of space-time processing of airborne radar data. Let n denote the sensor index, and let m be the time index. In the STAP problem, the target signal is modeled as a random amplitude complex exponential where the exponential is defined by a space-time steering vector that has the target's angle and Doppler. In other words, in the space-time domain the target model is that of a 2-D harmonic component similar to (6). The sum of the white noise field due to the internally generated receiver amplifier noise, and the colored noise field due to the sky noise contribution, is the purely indeterministic component of the space-time field decomposition. The presence of a jammer results in a barrage of noise localized in angle and uniformly distributed over all Doppler frequencies. Hence, in the space-time domain, each jammer is modeled as an evanescent component with $(a, b) = (0, 1)$ such that its 1-D modulating process $s_i^{(0,1)}(m)$ is the random process of the jammer amplitudes. The jammer samples from different pulses are uncorrelated. In the angle-Doppler domain each jammer contributes a 1-D delta function, parallel to the Doppler axis and located at a specific angle $\nu_i^{(0,1)}$ [using the notation of (5)]. The ground clutter results in an additional evanescent component of the observed 2-D space-time field. The clutter's echo from a single ground patch has a Doppler frequency that linearly depends on its aspect with respect to the platform. Hence, clutter from all angles lies in a "clutter ridge," which is supported on a diagonal line (that generally wraps around in Doppler) in the angle-Doppler domain. A model of the clutter field is then given by (5) with the slope of the clutter ridge given by b/a and with $s_i^{(a,b)}(na + mb)$ being a 1-D colored noise process. Since the rational numbers are dense in the set of real numbers, an irrational slope of the clutter ridge can be approximated arbitrarily close by a rational one. Hence, any clutter signal can be either exactly modeled or approximated by an evanescent field.

Fig. 2 graphically illustrates a typical example of the matching between the 2-D Wold decomposition based parametric random field model and the physical model of STAP data. In this synthetic example, the observed random field is the sum of two evanescent components that correspond to the clutter component with $(a, b) = (1, 2)$, $\nu^{(1,2)} = 0$ and a jammer with $\nu^{(0,1)} = 0.2$. Fig. 2 depicts the magnitude of the DFT of the observed field.

We therefore conclude that the foregoing derivation opens the way for new *parametric* solutions that can simplify and improve existing methods of STAP.

IV. Covariance Structure of the Observed Field

Based on the random field model derived in the previous sections, we derive in this section a closed-form parametric expression for the covariance matrix of the observed STAP data field in terms of the model parameters. We begin by stating our assumptions.

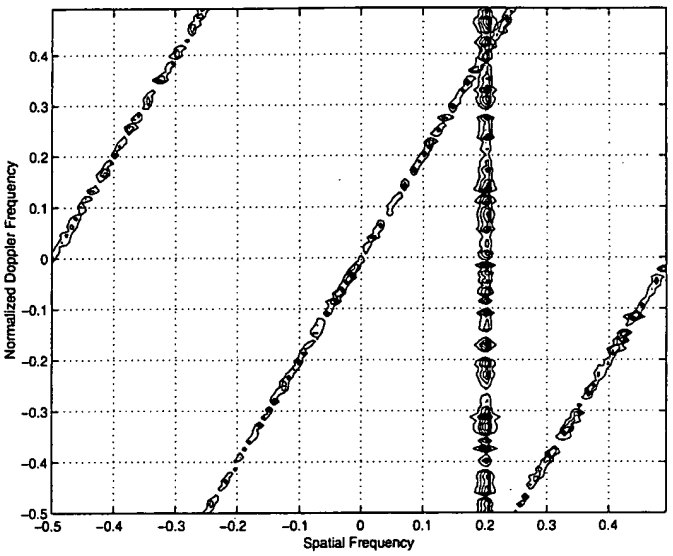


Fig. 2. Magnitude of the DFT of an observed field containing two evanescent components that correspond to a clutter component with $(a, b) = (1, 2)$, $\nu^{(1,2)} = 0$, and a jammer with $\nu^{(0,1)} = 0.2$.

Let $\{y(n, m)\}$, $(n, m) \in D$, where $D = \{(i, j)|0 \le i \le S-1, 0 \le j \le T-1\}$, be the observed random field.

Assumption 1: The purely indeterministic component $\{w(n, m)\}$ is a zero mean circular complex valued random field.

Assumption 2: The number $I = \sum_{(a,b)\in O} I^{(a,b)}$ of evanescent components in the field is *a priori* known. This assumption can be later relaxed.

Assumption 3: For each evanescent field $\{e_i^{(a,b)}\}$, the modulating 1-D purely indeterministic process $\{s_i^{(a,b)}\}$ is a zero-mean circular complex valued process.

Let

$$\mathbf{y} = [y(0,0), \ldots, y(0,T-1), y(1,0), \ldots, y(1,T-1) \ldots, \ldots, y(S-1,0), \ldots, y(S-1,T-1)]^T \quad (7)$$

$$\mathbf{w} = [w(0,0), \ldots, w(0,T-1), w(1,0), \ldots, w(1,T-1) \ldots, \ldots, w(S-1,0), \ldots, w(S-1,T-1)]^T \quad (8)$$

$$\mathbf{e}_i^{(a,b)} = \left[e_i^{(a,b)}(0,0), \ldots, e_i^{(a,b)}(0,T-1), e_i^{(a,b)}(1,0) \ldots, e_i^{(a,b)}(1,T-1), \ldots, \ldots, e_i^{(a,b)}(S-1,0) \ldots, e_i^{(a,b)}(S-1,T-1)\right]^T. \quad (9)$$

Let

$$\xi_i^{(a,b)} = \left[s_i^{(a,b)}(0), s_i^{(a,b)}(b), \ldots, s_i^{(a,b)}((T-1)b) \; s_i^{(a,b)}(a), s_i^{(a,b)}(a+b), \ldots, s_i^{(a,b)}(a+(T-1)b) \ldots, s_i^{(a,b)}((S-1)a), s_i^{(a,b)}((S-1)a+b) \ldots, s_i^{(a,b)}((S-1)a+(T-1)b)\right]^T \quad (10)$$

be the vector whose elements are the observed samples from the 1-D modulating process $\{s_i^{(a,b)}\}$. Define

$$\mathbf{v}^{(a,b)} = [0, d, \ldots, (T-1)d, c, c+d, \ldots \; c+(T-1)d, \ldots, \ldots, (S-1)c, (S-1)c+d \; \ldots, (S-1)c+(T-1)d]^T. \quad (11)$$

Given a scalar function $f(v)$, we will denote the matrix, or column vector, consisting of the values of $f(v)$ evaluated for all the elements of $\mathbf{v}$, where $\mathbf{v}$ is a matrix, or a column vector, by $f(\mathbf{v})$. Using this notation, we define

$$\mathbf{d}_i^{(a,b)} = \exp\left(j2\pi\nu_i^{(a,b)}\mathbf{v}^{(a,b)}\right). \quad (12)$$

Thus, using (5), we have that

$$\mathbf{e}_i^{(a,b)} = \xi_i^{(a,b)} \odot \mathbf{d}_i^{(a,b)} \quad (13)$$

where $\odot$ denotes an element-by-element product of the vectors.

Note that whenever $na + mb = ka + \ell b$ for some integers n, m, k, ℓ such that $0 \le n, k \le S-1$ and $0 \le m, \ell \le T-1$, the same sample from the modulating process $\{s_i^{(a,b)}\}$ is duplicated in the elements of $\xi_i^{(a,b)}$. It is shown in [15] that for a rectangular observed field of dimensions $S \times T$, the number of *distinct* samples from the random process $\{s_i^{(a,b)}\}$ that are found in the observed field is

$$N_c = (S-1)|a| + (T-1)|b| + 1 - (|a|-1)(|b|-1). \quad (14)$$

This is because N_c is the number of different "columns" one can define on such a rectangular lattice for a RNSHP defined by (a, b). We note here that in the special case where $a = 1$, (14) provides the well-known Brennan rule [3] on the rank of the clutter covariance matrix.

We therefore define the *concentrated version* $\mathbf{s}_i^{(a,b)}$ of $\xi_i^{(a,b)}$ to be an N_c-dimensional column vector of nonrepeating samples of the process $\{s_i^{(a,b)}\}$. More specifically, for the case in which $a > 0$ and $b < 0$, $\mathbf{s}_i^{(a,b)}$ is given by

$$\mathbf{s}_i^{(a,b)} = \left[s_i^{(a,b)}((T-1)b), \ldots, \ldots, s_i^{(a,b)}((S-1)a)\right]^T \quad (15)$$

whereas for the case in which $a \ge 0$ and $b \ge 0$, $\mathbf{s}_i^{(a,b)}$ is given by

$$\mathbf{s}_i^{(a,b)} = \left[s_i^{(a,b)}(0), \ldots \; \ldots, s_i^{(a,b)}((S-1)a + b(T-1))\right]^T. \quad (16)$$

Thus, for any (a, b), we have that

$$\xi_i^{(a,b)} = \mathbf{A}_i^{(a,b)}\mathbf{s}_i^{(a,b)} \quad (17)$$

where $\mathbf{A}_i^{(a,b)}$ is rectangular matrix of zeros and ones that replicates rows of $\mathbf{s}_i^{(a,b)}$.

Note, however, that due to boundary effects, the vector $\mathbf{s}_i^{(a,b)}$ is not composed of consecutive samples from the process $\{s_i^{(a,b)}\}$ unless $|a| \le 1$ or $|b| \le 1$. In other words, for some arbitrary a and b, there are missing samples in $\mathbf{s}_i^{(a,b)}$.

We note that the covariance matrix $\mathbf{R}_i^{(a,b)}$, which characterizes the second-order properties of the process $\{s_i^{(a,b)}\}$, is defined in terms of the concentrated version vector $\mathbf{s}_i^{(a,b)}$, i.e.,

$$\mathbf{R}_i^{(a,b)} = E\left[\mathbf{s}_i^{(a,b)}\left(\mathbf{s}_i^{(a,b)}\right)^H\right] \tag{18}$$

and not in terms of the covariance matrix

$$\tilde{\mathbf{R}}_i^{(a,b)} = E\left[\boldsymbol{\xi}_i^{(a,b)}\left(\boldsymbol{\xi}_i^{(a,b)}\right)^H\right] \tag{19}$$

of the vector $\boldsymbol{\xi}_i^{(a,b)}$. The matrix $\tilde{\mathbf{R}}_i^{(a,b)}$ is a singular matrix, where $\tilde{\mathbf{R}}_i^{(a,b)} = \mathbf{A}_i^{(a,b)}\mathbf{R}_i^{(a,b)}(\mathbf{A}_i^{(a,b)})^H$.

Since the evanescent components $\{e_i^{(a,b)}\}$ are mutually orthogonal and since all the evanescent components are orthogonal to the purely indeterministic component, we conclude that $\boldsymbol{\Gamma}$, which is the covariance matrix of $\mathbf{y}$, has the form

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_{\mathrm{PI}} + \sum_{(a,b)\in O}\sum_{i=1}^{I^{(a,b)}} \boldsymbol{\Gamma}_i^{(a,b)} \tag{20}$$

where $\boldsymbol{\Gamma}_i^{(a,b)}$ is the covariance matrix of $\mathbf{e}_i^{(a,b)}$.

Using (5) and (13), we find that

$$\boldsymbol{\Gamma}_i^{(a,b)} = \left(\mathbf{A}_i^{(a,b)}\mathbf{R}_i^{(a,b)}\left(\mathbf{A}_i^{(a,b)}\right)^T\right) \odot \left(\mathbf{d}_i^{(a,b)}\left(\mathbf{d}_i^{(a,b)}\right)^H\right). \tag{21}$$

A compact matrix representation of $\boldsymbol{\Gamma}_i^{(a,b)}$ for any (a, b) cannot be derived due to the dependence of the matrix structure on (a, b). However, for the case in which $(a, b) = (0, 1)$ (and similarly for $(a, b) = (1, 0)$), a somewhat more compact representation is possible, using Kronecker products instead of the Hadamard products.

More specifically, for this special case, (13) can be expressed in the form

$$\mathbf{e}_i^{(0,1)} = \mathbf{d}_i^{(0,1)} \otimes \mathbf{s}_i^{(0,1)} \tag{22}$$

where $\otimes$ is the Kronecker product. Hence

$$\boldsymbol{\Gamma}_i^{(0,1)} = \left(\mathbf{d}_i^{(0,1)}\left(\mathbf{d}_i^{(0,1)}\right)^H\right) \otimes \mathbf{R}_i^{(0,1)} = \mathbf{E}_i^{(0,1)} \otimes \mathbf{R}_i^{(0,1)} \tag{23}$$

where $\mathbf{R}_i^{(0,1)}$ and $\mathbf{E}_i^{(0,1)}$ are Toeplitz matrices, given by (24) and (25), as shown at the bottom of the page.

V. Parametric Estimation of the Interference Components

In this section, we derive a computationally efficient algorithm for estimating both the jamming and clutter fields, based on the above results. More specifically, for each interference component of the observed field, we estimate its spectral support parameters a, b, $\nu^{(a,b)}$ as well as c, d and the parametric model of the modulating 1-D purely indeterministic process $\{s_i^{(a,b)}\}$. In the setting of the radar problem considered here, partial information on the different components of the field is *a priori* known: The jamming signals are localized in angle and distributed over all Doppler frequencies. Thus, each jammer contributes an evanescent component with spectral support parameters $(a, b) = (0, 1)$ and an unknown frequency $\nu_i^{(0,1)}$. The clutter signal is also modeled as an evanescent component with $\nu^{(a,b)} = 0$ and an unknown (a, b) pair, which is uniquely determined by the platform motion parameters.

The proposed estimation algorithm of the spectral support parameters of the evanescent field a, b and $\nu_i^{(a,b)}$ is based on the following lemma.

Lemma 1: Let $\{e_i^{(a,b)}(n, m)\}$ be an evanescent field and let k be an integer. The samples of the evanescent field along a line

$$\mathbf{R}_i^{(0,1)} = \begin{bmatrix} r_i^{(0,1)}(0) & r_i^{(0,1)}(-1) & \cdots & r_i^{(0,1)}(-(T-1)) \\ r_i^{(0,1)}(1) & r_i^{(0,1)}(0) & \cdots & r_i^{(0,1)}(-(T-2)) \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & & r_i^{(0,1)}(-1) \\ r_i^{(0,1)}(T-1) & r_i^{(0,1)}(T-2) & \cdots & r_i^{(0,1)}(0) \end{bmatrix} \tag{24}$$

$$\mathbf{E}_i^{(0,1)} = \begin{bmatrix} 1 & \exp\left(-j2\pi\nu_i^{(0,1)}\right) & \cdots & \exp\left(-j2\pi(S-1)\nu_i^{(0,1)}\right) \\ \exp\left(j2\pi\nu_i^{(0,1)}\right) & 1 & \cdots & \exp\left(-j2\pi(S-2)\nu_i^{(0,1)}\right) \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & & \exp\left(-j2\pi\nu_i^{(0,1)}\right) \\ \exp\left(j2\pi(S-1)\nu_i^{(0,1)}\right) & \exp\left(j2\pi(S-2)\nu_i^{(0,1)}\right) & \cdots & 1 \end{bmatrix} \tag{25}$$

on the sampling grid defined by $k = na + mb$ are the samples of a 1-D constant amplitude harmonic signal, whose frequency is $\nu_i^{(a,b)}$.

Proof: Since for fixed a, b, k, $k = na + mb$ is the linear Diophantine equation (see the Appendix), its solutions are given by

$$n = n_k + tb \tag{26}$$

$$m = m_k - ta \tag{27}$$

where (n_k, m_k) is a solution of the equation, and t is an integer such that the sequence of consecutive values of t corresponds to the different lattice point on the line $k = na + mb$. From (5), we have, for the evanescent field samples along the line $k = na + mb$

$$\begin{aligned} e_i^{(a,b)}(n, m) &= s_i^{(a,b)}(na + mb) \exp\left(j2\pi\nu_i^{(a,b)}(nc + md)\right) \\ &= s_i^{(a,b)}(k) \exp\left(j2\pi\nu_i^{(a,b)}(n_k c + m_k d - t(ad - bc))\right) \\ &= \left[s_i^{(a,b)}(k) \exp\left(j2\pi\nu_i^{(a,b)}(n_k c + m_k d)\right)\right] \\ &\quad \cdot \exp\left(-j2\pi\nu_i^{(a,b)} t\right) \end{aligned} \tag{28}$$

where the last equality is because c, d are coprime integers such that $ad - bc = 1$. Hence, in each realization and for a fixed k, $s_i^{(a,b)}(k) \exp(j2\pi\nu_i^{(a,b)}(n_k c + m_k d))$ is a (random) constant. Hence, the proof follows. ■

The algorithm is implemented by the following four-step procedure:

Initial estimation of a and b: In the presence of an evanescent component, the peaks of the observed field periodogram are concentrated along a straight line such that its slope is defined by the two coprime integers a and b. Hence, several alternative approaches for obtaining an initial estimate of the spectral support parameters of the evanescent component can be derived by taking the Radon or Hough transforms [20] of the observed field periodogram. (The current implementation employs the Hough transform for detecting straight lines in 2-D arrays). However, due to noise presence, this estimate may perturb. Since, on a finite-dimension observed field, only a finite number of possible (a, b) pairs may be defined, the output of the initial stage is a set of possible (a, b) pairs such that the ratio b/a is close to the ratio obtained for the (a, b) pair estimated by the Hough transform.

Estimation of the frequency parameter of the evanescent component: For each possible (a, b) pair, we next evaluate the frequency parameter of the evanescent component $\nu_i^{(a,b)}$. Assuming the considered (a, b) pair is the correct one, we have, from Lemma 1, that in the absence of background noise for a fixed $k = na + mb$ (i.e., along a line on the sampling grid), the samples of the evanescent component are the samples of a 1-D constant amplitude harmonic signal, whose frequency is $\nu_i^{(a,b)}$. Hence, by considering the samples along such a line, we obtain samples of a 1-D constant amplitude harmonic signal whose frequency $\nu_i^{(a,b)}$ can be easily estimated using any standard frequency estimation algorithm (e.g., the 1-D DFT).

Final estimation of the spectral support parameters of each evanescent component: The test for detecting the correct (a, b) and $\nu_i^{(a,b)}$ is then based on multiplying the observed signal $y(n, m)$ by $\exp(-j2\pi\hat{\nu}_i^{(\hat{a},\hat{b})}(n\hat{c} + m\hat{d}))$ for each of the considered $\hat{a}$, $\hat{b}$ and $\hat{\nu}_i^{(\hat{a},\hat{b})}$ triplets and evaluating the variance of this signal along a line on the sampling grid such that $k = n\hat{a} + m\hat{b}$. Clearly, the best estimate of a, b, and $\nu_i^{(a,b)}$ is the one that results in minimal variance for the 1-D sequence because in the absence of noise, the correct a, b, and $\nu_i^{(a,b)}$ result in a zero variance. Note that $\hat{c}$, $\hat{d}$ are two coprime integers satisfying the linear Diophantine equation $ad - bc = 1$ when a, b are replaced by their estimated values. Clearly, $\hat{c}$, $\hat{d}$ obtained as solutions to the linear Diophantine equation are not unique (see the Appendix). The correct pair $\hat{c}$, $\hat{d}$ is then determined by employing the symmetry properties of the field covariance sequence (see [12] for details). Since, in the STAP problem, it is *a priori* known that for the jammers $(a, b) = (0, 1)$, whereas for the clutter $\nu^{(a,b)} = 0$, the parameters c, d do not appear in the model and, hence, need not be estimated. Nevertheless, to maintain the generality of the algorithm description, we proceed for the final step of the algorithm with the general description, assuming $\hat{c}$, $\hat{d}$ have been estimated (or are *a-priori* known as in the STAP case).

Estimating the model of the 1-D purely indeterministic modulating process of the evanescent field: Having estimated the spectral support parameters of each evanescent component, we take the approach of first estimating a *nonparametric* representation of its 1-D purely indeterministic modulating process $\{s_i^{(a,b)}\}$, and only at a second stage do we estimate the parametric models of these processes. Hence, in the first stage, we estimate the particular values that the vectors $\xi_i^{(a,b)}$ take for the given realization, i.e., we treat these as unknown constants. The estimation procedure is implemented as follows: Multiplying the observed signal $y(n, m)$ by $\exp(-j2\pi\hat{\nu}_i^{(\hat{a},\hat{b})}(n\hat{c} + m\hat{d}))$ and evaluating the arithmetic mean of this signal along a line on the sampling grid such that $k = n\hat{a} + m\hat{b}$, we have

$$\hat{s}_i^{(a,b)}(k) = \frac{1}{N_s} \sum_{n\hat{a}+m\hat{b}=k} y(n, m) \cdot \exp(-j2\pi\hat{\nu}_i^{(a,b)}(n\hat{c} + m\hat{d})) \tag{29}$$

where N_s denotes the number of the observed field samples that satisfy the relation $n\hat{a}+m\hat{b} = k$. Once we obtained the sequence of estimated samples from the 1-D modulating process $\{s_i^{(a,b)}\}$, the problem of estimating its parametric model becomes entirely a 1-D estimation problem. Assuming the modulating process is an autoregressive (AR) process and applying to the sequence an AR estimation algorithm (see, e.g., [21]), we obtain estimates of the modulating process parameters as well.

Finally, it is important to note that we solve the difficult problem of evaluating the rank of the low-rank covariance matrix of the interference as a byproduct of obtaining the parametric estimates of the interference components: Denote the number of evanescent components (interference sources) of the field by Q. It is then shown in [16] that the rank of the interference covariance matrix is given by $S\sum_{k=1}^{Q}|a_k| + T\sum_{k=1}^{Q}|b_k| - \sum_{k=1}^{Q}|a_k|\sum_{k=1}^{Q}|b_k|$. In fact,

the special case where $Q = 1$ and $a = 1$ is the well-known Brennan rule [3] on the rank of the clutter covariance matrix. Hence, following the estimation of the spectral support parameters of the different evanescent components, the rank of the interference covariance matrix is also determined.

VI. Parametric Fully Adaptive Processing

Having estimated the parametric models of the purely indeterministic and evanescent components of the field, the estimated parameters can be substituted into (20) and (21) to obtain an estimate of the interference-plus-noise covariance matrix $\mathbf{\Gamma}$. In this section, we show how the estimated interference-plus-noise covariance matrix is employed to obtain a fully adaptive space-time filter.

Let $\mathbf{v}_t$ denote the target steering vector given by

$$\mathbf{v}_t = \mathbf{b}(\varpi_t) \otimes \mathbf{a}(\vartheta_t). \tag{30}$$

Assuming a linear, uniformly spaced, sensor array and a uniform coherent-processing interval (CPI) are employed in our model, the spatial steering vector $\mathbf{a}(\vartheta)$ and the temporal steering vector $\mathbf{b}(\varpi)$ are given by

$$\mathbf{a}(\vartheta) = [1,\, e^{j2\pi\vartheta}, \ldots, e^{j2\pi(T-1)\vartheta}]^T$$

$$\mathbf{b}(\varpi) = [1,\, e^{j2\pi\varpi}, \ldots, e^{j2\pi(S-1)\varpi}]^T$$

respectively. Assume for the moment that only a single target may exist in the observed data and that both the target's steering vector and the interference-plus-noise covariance matrix $\mathbf{\Gamma}$ are known. We next derive a fully adaptive detection algorithm based on the generalized likelihood ratio test (GLRT). Since $\mathbf{v}_t$ and $\mathbf{\Gamma}$ are assumed known, the GLR has to be maximized only with respect to C_t, which is the unknown amplitude parameter of the target. Thus, the GLR has the form

$$\Lambda = \frac{\max_{C_t} p_{\mathbf{y}|\mathcal{H}_1}(\mathbf{y};\, C_t|\mathcal{H}_1)}{p_{\mathbf{y}|\mathcal{H}_0}(\mathbf{y}|\mathcal{H}_0)}. \tag{31}$$

Following a standard procedure (see, e.g., [7] and [9]), the GLR test statistic, which we denote by $|z(\varpi,\, \vartheta)|^2$, can be shown to have the equivalent form

$$|z(\varpi_t,\, \vartheta_t)|^2 = \frac{|\mathbf{v}_t^H \mathbf{\Gamma}^{-1}\mathbf{y}|^2}{\mathbf{v}_t^H \mathbf{\Gamma}^{-1}\mathbf{v}_t}. \tag{32}$$

Let $\mathbf{\Psi} = \mathbf{\Gamma}^{-1}\mathbf{y}$. We thus have

$$|z(\varpi_t,\, \vartheta_t)|^2 = \frac{|\mathbf{v}_t^H \mathbf{\Psi}|^2}{\mathbf{v}_t^H \mathbf{\Gamma}^{-1}\mathbf{v}_t} = \frac{|\mathbf{b}^H(\varpi_t) \otimes \mathbf{a}^H(\vartheta_t)\mathbf{\Psi}|^2}{\mathbf{v}_t^H \mathbf{\Gamma}^{-1}\mathbf{v}_t}. \tag{33}$$

Reorganizing the elements of $\mathbf{\Psi}$ into a $S \times T$ matrix $\mathbf{\Upsilon}$ where the elements of the kth row of $\mathbf{\Upsilon}$ are $\mathbf{\Psi}((k-1)T+1) \cdots \mathbf{\Psi}(kT)$, we conclude that for a linear, uniformly spaced, sensor array and uniform CPI

$$\mathbf{b}^H(\varpi) \otimes \mathbf{a}^H(\vartheta)\mathbf{\Psi} = \sum_{p=1}^{S}\sum_{q=1}^{T} e^{-j2\pi(p-1)\varpi} e^{-j2\pi(q-1)\vartheta} \mathbf{\Upsilon}(p,\, q). \tag{34}$$

Thus, $\mathbf{b}^H(\varpi) \otimes \mathbf{a}^H(\vartheta)\mathbf{\Psi}$ and $\mathbf{\Upsilon}$ are a 2-D DFT pair. However, since in fact the steering vector is unknown, the detector must first estimate the frequency where the magnitude of the 2-D DFT of $\mathbf{\Upsilon}$ is maximal, followed by comparison of the value of the test statistic evaluated at this frequency against the threshold. Thus, the GLRT when $\mathbf{\Gamma}$ is perfectly known is given by

$$\max_{\varpi,\vartheta} |z(\varpi,\, \vartheta)|^2 \gtrless \gamma. \tag{35}$$

In other words, in the case of a known covariance matrix, the test is equivalent to finding the 2-D frequency where the magnitude of the 2-D DFT of $\mathbf{\Upsilon}$ is maximal, followed by comparison of the value of the test statistic at this frequency against the threshold.

Note that under both the null hypothesis (no target) $\mathcal{H}_0$ as well as under the alternative hypothesis $\mathcal{H}_1$, $\mathbf{b}^H(\varpi) \otimes \mathbf{a}^H(\vartheta)\mathbf{\Psi}$ is a Gaussian random variable, being a linear transformation of a Gaussian random vector. Assuming $\mathbf{\Gamma}$ is perfectly known, it is not difficult to show [13] that after prewhitening by $\mathbf{\Gamma}^{-(1/2)}$, the probability density function of the GLRT in (35) is χ^2 distributed with two degrees of freedom under $\mathcal{H}_0$ and noncentral χ^2 with two degrees of freedom under $\mathcal{H}_1$.

Finally, since $\mathbf{\Gamma}$ is also unknown, we adopt an approach similar to that employed in the derivation of the adaptive match filter (AMF) in [7] and substitute the unknown covariance matrix with its estimate, which is obtained as explained in the previous sections.

To illustrate the operation of the proposed solution, we resort to numerical evaluation of some specific examples (see [13] for a detailed performance analysis and additional examples). Consider a 2-D observed random field consisting of a sum of a purely indeterministic component (background noise), a single evanescent (interference) component, and three harmonic components (targets). The purely indeterministic component is a complex valued circular Gaussian white noise field. The evanescent component spectral support parameters are $(a,\, b) = (1,\, 2)$, $\nu^{(1,2)} = 0$. The modulating 1-D purely indeterministic process of this evanescent component is a first-order Gaussian AR process, with driving noise variance $(\sigma^{(1,2)})^2 = 2$ and $a^{(1,2)}(1) = -0.5$. There are three targets that are located at $(0.05,\, 0)$, $(0.15,\, 0.15)$, and $(-0.25,\, 0.15)$, respectively. The observed field dimensions are 48×48.

Let us define the power of each of the field components as $E_w = \mathbf{w}^H\mathbf{w}$ for the purely indeterministic component; $E_e = (\mathbf{e}^{(a,b)})^H\mathbf{e}^{(a,b)}$ for the evanescent component; and $E_{h_k} = \mathbf{h}_k^H\mathbf{h}_k$, $k = 1,\, 2,\, 3$ for each of the harmonic components, where $\mathbf{h}_k$ is defined in the same way $\mathbf{w}$ and $\mathbf{e}^{(a,b)}$ are defined. In this example, we have $E_e/E_w = 6\,\text{dB}$, whereas for the three targets, we have $E_{h_1}/E_w = -12.8\,\text{dB}$, $E_{h_2}/E_w = -14.5\,\text{dB}$, $E_{h_3}/E_w = -15\,\text{dB}$. Due to the strong interference component, the presence of the three targets is hard to detect in the observed data whose power spectral density is depicted in Fig. 3. However, these targets are easily detected by the test statistic $|z(\varpi,\, \vartheta)|$, depicted in Fig. 4. In Fig. 4, $|z(\varpi,\, \vartheta)|$ is depicted as a function of the 2-D frequencies, i.e., angle and Doppler.

VII. Parametric Partially Adaptive Processing

The low rank of the interference covariance matrix is exploited in the partially adaptive STAP to significantly reduce the adaptive problem dimensionality. In this section, we derive a partially adaptive processing algorithm, based on the estimated parametric model of the interference. Moreover, it is proved in

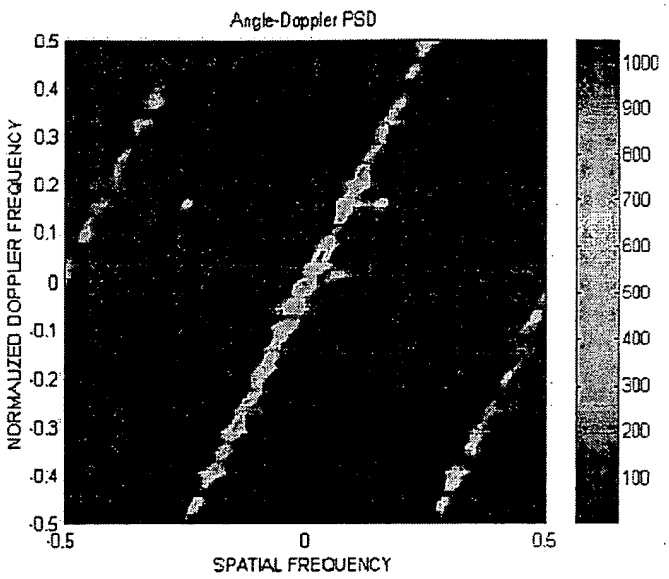


Fig. 3. Power spectral density of the observed field.

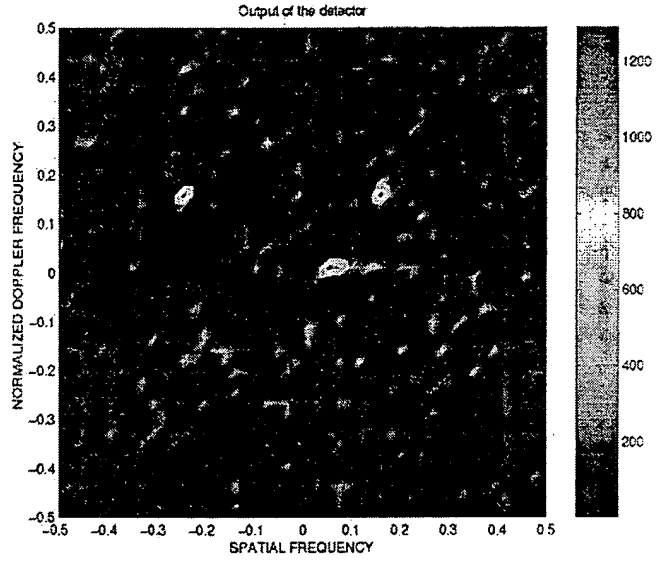


Fig. 4. Test statistic $|z(\varpi, \vartheta)|$.

this section that in order to implement the proposed partially adaptive processing method, *only* the spectral support parameters of the interference need to be estimated, and there is no need whatsoever to estimate the modulating process of the interference model, nor the data covariance matrix.

More specifically, recall that

$$\mathbf{\Gamma}_i^{(a,b)} = \left(\mathbf{A}_i^{(a,b)}\mathbf{R}_i^{(a,b)}\left(\mathbf{A}_i^{(a,b)}\right)^T\right) \odot \left(\mathbf{d}_i^{(a,b)}\left(\mathbf{d}_i^{(a,b)}\right)^H\right). \quad (36)$$

Having estimated a, b and $\nu_i^{(a,b)}$ using the algorithm in Section V, the vector $\mathbf{d}_i^{(a,b)}$ is known. Hence, demodulating $\mathbf{e}_i^{(a,b)}$, we conclude using (13) that

$$\xi_i^{(a,b)} = \mathbf{e}_i^{(a,b)} \odot \left((\mathbf{d}_i^{(a,b)})^H\right)^T. \quad (37)$$

However, from (17), we conclude that the covariance matrix of $\xi_i^{(a,b)}$ is given by

$$\tilde{\mathbf{R}}_i^{(a,b)} = \mathbf{A}_i^{(a,b)}\mathbf{R}_i^{(a,b)}\left(\mathbf{A}_i^{(a,b)}\right)^T. \quad (38)$$

In the following, we prove that since a and b are already known, an orthogonal projection matrix onto the low-rank subspace spanned by the evanescent field covariance matrix can be found *without* estimating the parametric model of the evanescent field 1-D modulating process and, hence, without estimating $\mathbf{R}_i^{(a,b)}$. Moreover, this result enables us to avoid the need in both evaluating the field covariance matrix and in employing a computationally intensive eigenanalysis to the estimated covariance matrix. More specifically, let us construct the following orthogonal projection matrix:

$$\mathbf{T}_i^{(a,b)} = \mathbf{A}_i^{(a,b)}\left(\left(\mathbf{A}_i^{(a,b)}\right)^T\mathbf{A}_i^{(a,b)}\right)^{-1}\left(\mathbf{A}_i^{(a,b)}\right)^T. \quad (39)$$

It is easily verified (by substitution) that $\mathbf{T}_i^{(a,b)}$ is an orthogonal projection onto the range space of $\tilde{\mathbf{R}}_i^{(a,b)}$ since for any ST-dimensional vector $\mathbf{v}$

$$\tilde{\mathbf{R}}_i^{(a,b)}\mathbf{v} = \tilde{\mathbf{R}}_i^{(a,b)}\mathbf{T}_i^{(a,b)}\mathbf{v} = \mathbf{T}_i^{(a,b)}\tilde{\mathbf{R}}_i^{(a,b)}\mathbf{v}. \quad (40)$$

In addition, $(\mathbf{T}_i^{(a,b)})^2 = \mathbf{T}_i^{(a,b)}$, and $(\mathbf{T}_i^{(a,b)})^T = \mathbf{T}_i^{(a,b)}$.

Note that since $\mathbf{A}_i^{(a,b)}$ is a sparse matrix of zeros and ones *only*, the computation of $\mathbf{T}_i^{(a,b)}$ is very simple. The projection matrix onto the subspace orthogonal to the interference space is therefore given by $(\mathbf{T}_i^{(a,b)})^\perp = \mathbf{I} - \mathbf{T}_i^{(a,b)}$. Hence, by projecting the demodulated observed data vector $\bar{\mathbf{y}} = \mathbf{y} \odot ((\mathbf{d}_i^{(a,b)})^H)^T$ onto the subspace orthogonal to the interference subspace, a reduced-dimension data vector given by $\tilde{\mathbf{y}} = (\mathbf{T}_i^{(a,b)})^\perp\bar{\mathbf{y}}$ is obtained, such that the interference contribution to the observed signal is mitigated. Remodulating $\tilde{\mathbf{y}}$ by evaluating $\tilde{\mathbf{y}} \odot \mathbf{d}_i^{(a,b)}$, followed by sequentially applying this procedure to mitigate each of the interference sources, the detection problem is reduced to that of detecting a target in the presence of background noise only. Following a similar derivation to the one in (31)–(35), we conclude that in the special case where the background noise is known to be a white noise field, the statistical test is obtained by finding the 2-D frequency where the magnitude of the 2-D DFT of the processed data vector (organized back into a 2-D array) is maximal, followed by comparison of the value of the test statistic at this frequency against the threshold. In the more general case, where the purely indeterministic component of the field is not a white noise field, the observed data vector is first prewhitened by the estimated $\Gamma_{PI}^{-(1/2)}$. It is shown in [13] that the probability density function of the GLR test that upper bounds the performance of the actual detector is χ^2 with two degrees of freedom under $\mathcal{H}_0$ and noncentral χ^2 with two degrees of freedom under $\mathcal{H}_1$.

As an example, consider the same field as in the previous section. Due to the strong interference component, the presence of the three targets is difficult to detect in the observed data, whose power spectral density is depicted in Fig. 3. However, these targets are easily detected in the processed data, as illustrated in Fig. 5. This result is obtained without estimating the parametric model of the evanescent field 1-D modulating process

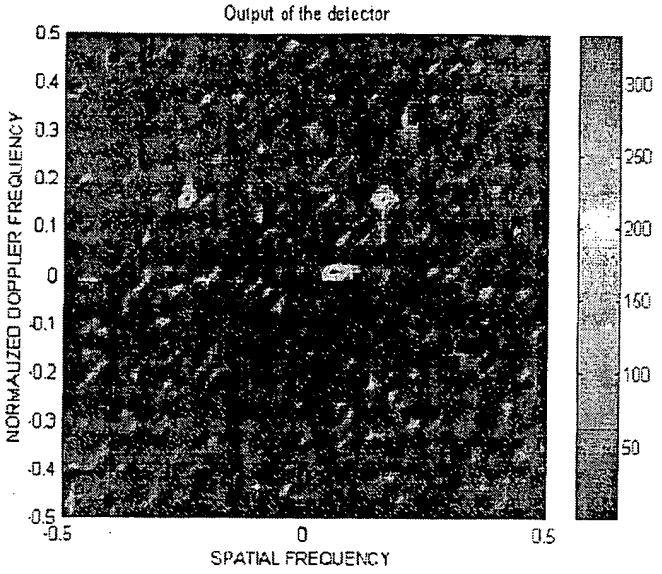


Fig. 5. Test statistic of the parametric partially adaptive processor. The power spectral density of the field after being projected onto the subspace orthogonal to the interference subspace.

and, hence, without estimating the interference-plus-noise covariance matrix. Since both the estimation of the interference-plus-noise covariance matrix, as well as its analysis, are saved, the proposed parametric partially adaptive processing method is robust and computationally attractive (see [13] for a detailed performance analysis and additional examples).

VIII. Conclusions

In this paper, a novel parametric approach for modeling, estimation, and target detection for STAP data has been derived. The proposed parametric interference mitigation procedures employ the information in only a single range gate, thus achieving high performance gain when the data in the different range gates cannot be assumed stationary. The model is based on the results of the 2-D Wold-like decomposition. We showed that the same parametric model that results from the 2-D Wold-like orthogonal decomposition naturally arises as the physical model in the problem of space-time processing of airborne radar data. We exploited this correspondence to derive computationally efficient fully adaptive and partially adaptive detection algorithms. Having estimated the models of the noise and interference components of the field, the estimated parameters are substituted into the parametric expression of the covariance matrix to obtain an estimate of the interference-plus-noise covariance matrix. Hence, the fully adaptive weight vector is obtained, and a corresponding test is derived. Moreover, we proved that it is sufficient to estimate only the spectral support parameters of each interference component in order to obtain a projection matrix onto the subspace orthogonal to the interference subspace. Thus, the resulting detector is statistically superior to the fully adaptive detector as considerably fewer parameters need to be estimated. Since a much smaller number of parameters need to be estimated the proposed partially adaptive detector is also computationally much simpler. Statistical analysis of the performance of the proposed detectors is considered in [13].

Appendix
Linear Diophantine Equation

Let k and ℓ be two nonzero integers and p some other integer. The equation

$$kx - \ell y = p$$

is called the *linear Diophantine equation*. A *solution* of this equation is a pair (x, y) of integers (a *lattice point* in the plane) that satisfies the equation. We use the following well known theorem (e.g., see [23])

Theorem 1: The linear Diophantine equation

$$kx - \ell y = p$$

has a solution if and only if $q \mid p$, (i.e., q divides p), where $q = g.c.d.(k, \ell)$. Furthermore, if (x_0, y_0) is a solution of this equation, then the set of solutions of the equation consists of all integer pairs (x, y) of the form

$$x = x_0 + t\frac{\ell}{q} \quad \text{and} \quad y = y_0 + t\frac{k}{q}, \quad t \in \mathcal{Z}. \tag{41}$$

Note that if k and ℓ are coprime, then there will always be solutions, given by (41).

Acknowledgment

The authors would like to thank W. Fu for his help in producing the numerical examples.

References

[1] *Proc. 1st IEEE Sensor Array Multichannel Signal Processing Workshop*, Cambridge, MA, Mar. 2000.
[2] *Proc. 8th Adaptive Sensor Array Processing Workshop*, Lexington, MA, Mar. 2000.
[3] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Lab., Mass. Inst. Technol., Lexington, MA, Tech. Rep. 1015, Dec. 1994.
[4] R. Klemm, *Space Time Adaptive Processing*. London, U.K.: IEE, 1998.
[5] R. Klemm, "Prospectives in STAP research," in *Proc. 8th Adaptive Sensor Array Processing Workshop*, Lexington, MA, Mar. 2000.
[6] A. Haimovich, "The Eigencanceler: Adaptive radar by eigenanalysis methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, pp. 532–542, Apr. 1996.
[7] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, pp. 208–216, Jan. 1992.
[8] Y. L. Gau and I. S. Reed, "An improved reduced-rank CFAR space-time adaptive radar detection algorithm," *IEEE Trans. Signal Process.*, vol. 46, pp. 2139–2146, Aug. 1998.
[9] J. S. Goldstein, I. S. Reed, and P. A. Zulch, "Multistage partially adaptive STAP CFAR detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, pp. 645–661, Apr. 1999.
[10] J. M. Francos, A. Z. Meiri, and B. Porat, "A Wold-like decomposition of 2-D discrete homogeneous random fields," *Ann. Appl. Prob.*, vol. 5, pp. 248–260, 1995.
[11] C. Cuny, "On the prediction of vector-valued random fields and the properties of their evanescent components,", submitted for publication.
[12] G. Cohen and J. M. Francos, "Spectral representation and asymptotic properties of certain deterministic fields with innovation components," *Prob. Theory Related Fields*, to be published.
[13] J. M. Francos and A. Nehorai, "Performance of the interference mitigation for STAP using the two-dimensional Wold decomposition model,", submitted for publication.

[14] J. M. Francos, "An exact Cramer–Rao bound on the estimation of complex valued homogeneous random fields," *IEEE Trans. Signal Process.*, vol. 50, pp. 710–724, Mar. 2002.
[15] ——, "Bounds on the accuracy of estimating the parameters of discrete homogeneous random fields with mixed spectral distributions," *IEEE Trans. Inform. Theory.*, vol. 43, pp. 908–922, May 1997.
[16] G. Cohen and J. M. Francos, "Efficient parameter estimation of evanescent random fields,", submitted for publication.
[17] T. K. Sarkar, J. Koh, R. Adve, R. A. Schneible, M. C. Wicks, S. Choi, and M. S. Palma, "A pragmatic approach to adaptive antennas," *IEEE Antennas Propagat. Mag.*, vol. 42, pp. 39–55, Apr. 2000.
[18] H. Wang and L. Cai, "On adaptive spatial-temporal processing for airborne surveillance radar systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, pp. 660–669, July 1994.
[19] X. Lin and R. S. Blum, "Robust STAP algorithms using prior knowledge for airborne radar applications," *Signal Process.*, vol. 79, pp. 273–287, 1999.
[20] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[21] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[22] S. M. Kay, *Fundamentals of Statistical Signal Processing—Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[23] G. E. Andrews, *Number Theory*. Philadelphia, PA: W. B. Saunders, 1971.

Joseph M. Francos (SM'97) was born on November 6, 1959, in Tel-Aviv, Israel. He received the B.Sc. degree in computer engineering in 1982 and the D.Sc. degree in electrical engineering in 1991, both from the Technion—Israel Institute of Technology, Haifa.

From 1982 to 1987, he was with the Signal Corps Research Laboratories, Israeli Defense Forces. From 1991 to 1992, he was with the Department of Electrical Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, as a Visiting Assistant Professor. During 1993, he was with Signal Processing Technology, Palo Alto, CA. In 1993, he joined the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva, Israel, where he is now an Associate Professor. He also held visiting positions at the Massachusetts Institute of Technology Media Laboratory, Cambridge; at the Electrical and Computer Engineering Department, University of California, Davis; at the Electrical Engineering and Computer Science Department, University of Illinois, Chicago; and at the Electrical Engineering Department, University of California, Santa Cruz. His current research interests are in parametric modeling and estimation of 2-D random fields, random fields theory, parametric modeling and estimation of nonstationary signals, space-time coding, image modeling and indexing, and texture analysis and synthesis.

Dr. Francos served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2001.

Arye Nehorai (S'80–M'83–SM'90–F'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, in 1976 and 1979, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1983.

After graduation, he worked as a Research Engineer for Systems Control Technology, Inc., Palo Alto, CA. From 1985 to 1995, he was with the Department of Electrical Engineering, Yale University, New Haven, CT, where he became an Associate Professor in 1989. In 1995, he joined the Department of Electrical Engineering and Computer Science at The University of Illinois at Chicago (UIC) as a Full Professor. From 2000 to 2001, he was Chair of the Department's Electrical and Computer Engineering (ECE) Division, which is now a new department. He holds a joint professorship with the ECE and Bioengineering Departments at UIC. His research interests are in signal processing, communications, and biomedicine.

Dr. Nehorai is Vice President-Publications of the IEEE Signal Processing Society. He was Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from January 2000 to December 2002. He is currently a Member of the Editorial Board of *Signal Processing*, the IEEE SIGNAL PROCESSING MAGAZINE, and *The Journal of the Franklin Institute*. He has previously been an Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, the IEEE JOURNAL OF OCEANIC ENGINEERING, and *Circuits, Systems, and Signal Processing*. He served as Chairman of the Connecticut IEEE Signal Processing Chapter from 1986 to 1995 and as a Founding Member, Vice-Chair, and later Chair of the IEEE Signal Processing Society's Technical Committee on Sensor Array and Multichannel (SAM) Processing from 1998 to 2002. He was the co-General Chair of the First and Second *IEEE SAM Signal Processing Workshops* held in 2000 and 2002. He was co-recipient, with P. Stoica, of the 1989 IEEE Signal Processing Society's Senior Award for Best Paper. He received the Faculty Research Award from the UIC College of Engineering in 1999 and was Adviser of the UIC Outstanding Ph.D. Thesis Award in 2001. In 2001, he was named University Scholar of the University of Illinois. He has been a Fellow of the Royal Statistical Society since 1996.

Subspace Approximation for Adaptive Multichannel Radar Filtering

A.W. Bojanczyk
School of Electrical Engineering
Cornell University
Ithaca, NY
adamb@ee.cornell.edu

W.L. Melvin and E.J. Holder
Georgia Tech Research Institute/SEAL
Georgia Institute of Technology
Atlanta, GA
bill.melvin@gtri.gatech.edu or jeff.holder

Abstract

In this paper we consider subspace approximation tailored to adaptive airborne radar. Motivation for this research includes the need for reduced computational burden and approaches for practical implementation. Measured radar data only approximately satisfies the statistical assumptions intrinsic to the adaptive processor. Hence, approximate numerical methods for adaptive weight computation may successfully be used in place of exact methods. We propose a numerical procedure based on partial bi-diagonalization of the interference covariance matrix, coupled with a pre-conditioned conjugate gradient iterative method, to extract approximate basis vectors for the interference subspace. We use these basis vectors to construct an adaptive weight vector. Through example, we show the potential of this method for adaptive radar.

1. Introduction

Airborne radar must detect targets of diminishing radar cross-section, often at decreased radial velocities. The interference environment is severe and poses a significant challenge to effective target detection. Exploiting signal diversity over multiple domains offers enhanced detection performance. Space-time adaptive processing (STAP) represents a class of multi-domain adaptive techniques useful in such circumstances. The theory of adaptive radar was proposed in a series of papers by Brennan, Mallett and Reed in the early 1970's [1-2]. STAP remains an active research topic and most view advanced STAP techniques as a vital component of future airborne and spaceborne radar systems.

Recent STAP research focuses largely on mitigating computational burden and/or proposing novel processing architectures [3-8]. Eigenbased STAP, also called reduced-rank or partially adaptive STAP methods, rely on a sequence of linear transformations and selection operators to generate a set of adaptive weights [5-8]. Such methods approach optimum performance, but at the expense of considerably increased computational complexity. This occurs because the sequence of required operations amounts to a singular value decomposition (SVD) of the data matrix. Nevertheless, such approaches leverage the numerical stability of the SVD and require minimal sample support for covariance matrix estimation. These considerations are important when applying STAP in routinely encountered inhomogeneous and sample-poor signal environments.

In this paper we propose approximate numerical techniques as an alternative approach for the airborne radar STAP problem. The proposed approach capitalizes on: (1) the generally low numerical rank of the covariance matrix; (2) the realization that actual airborne radar signal environments only approximately satisfy assumed statistical models; and, (3) low-rank realizations require smaller sample populations for parameter estimation.

2. Adaptive Airborne Radar

Consider an M channel, aircraft-mounted array receiving N pulses. The space-time snapshot, $x_k \in C^{MN}$, is given by

$$X_k = [x_{(k,1,1)}, \cdots, x_{(k,M,1)}, x_{(k,1,2)}, \cdots, x_{(k,M,2)}, \cdots, x_{(k,1,N)}, \cdots, x_{(k,M,N)}]^T \quad (1)$$

with $x_{(k,m,n)}$ representing a complex baseband observation at the k[th] range, m[th] channel and n[th] pulse. Each snapshot consists of additive signal contributions from clutter, $X_{k,C}$, jamming, $X_{k,J}$, uncorrelated noise, $X_{k,N}$, and targets, $X_{k,T}$, such that

$$X_k = \begin{cases} X_{k|H_0} = X_{k,C} + X_{k,J} + X_{k,N} \\ X_{k|H_1} = X_{k,C} + X_{k,J} + X_{k,N} + X_{k,T} \end{cases} \quad (2)$$

for the null and alternative hypotheses. The output of the adaptive processor is given by

$$y_k = \hat{W}_k^H X_k , \quad (3)$$

where $\hat{W}_k \in C^{MN}$ is the adaptive weight vector. This weight vector takes the general form

$$\hat{W}_k = \hat{\alpha}_k \hat{R}_k^{-1} S_T \; ; \quad \hat{R}_k = \frac{1}{\kappa} \sum_{\substack{m=1 \\ m \neq k}}^{\kappa} X_m X_m^H \; , \qquad (4)$$

where $S_T \in C^{MN}$ is the target space-time steering vector, $\hat{R}_k$ is the maximum likelihood estimate (MLE) of the interference covariance matrix [2], and $\hat{\alpha}_k$ is a constant. The space-time steering vector represents the response of the array to a point source with a specific direction of arrival and Doppler frequency. Note that $\hat{R}_k$ approximates $E[X_{k/H_0} X_{k/H_0}^H]$. We subsequently subject y_k to binary hypothesis testing as a means of declaring target presence.

3. Eigenbased Adaptive Radar

The sample covariance matrix $\hat{R}_k$ is Hermitian and positive definite. For this reason, a unitary matrix Q_k exists such that

$$\hat{R}_k = Q_k \Lambda_k Q_k^H = \sum_{m=1}^{MN} \lambda_k(m) \, q_k(m) \, q_k^H(m) \; , \qquad (5)$$

where $\lambda_k(1) \geq \lambda_k(2) \geq \ldots \geq \lambda_k(MN)$, $q_k(m)$ is the mth column of Q_k and I_{MN} is the MN x MN identity matrix. $\lambda_k(m)$ and $q_k(m)$ are the mth eigenvalue and eigenvector of $\hat{R}_k$, respectively. Decomposing (5) into principal components (PC) and noise yields

$$\begin{aligned} Q_{PC,k} &= \left[q_k(1), q_k(2), \ldots, q_k(P) \right] \\ \lambda_k(1) &\geq \lambda_k(2) \geq \ldots \geq \lambda_k(P) \; , \end{aligned} \qquad (6)$$

and

$$\begin{aligned} Q_{noise,k} &= \left[q_k(P+1), q_k(P+2), \ldots, q_k(MN) \right] \\ \lambda_k(P) &\gg \lambda_k(P+1) \approx \lambda_k(P+2) \approx \ldots \approx \lambda_k(MN) \; . \end{aligned} \qquad (7)$$

Observe that $Q_{PC,k} \perp Q_{noise,k}$. The principal and noise components define the interference and noise subspaces.

An interesting result in [5] shows that explicit knowledge of the true interference subspace completely solves the optimum filtering problem. Assuming matched channels and letting $\hat{\alpha}_k \rightarrow \alpha_k = 1$, we may write (4) as

$$W_k = \frac{1}{\lambda_0} \left[S_T - \sum_{m=1}^{NM} \frac{\lambda_k(m) - \lambda_0}{\lambda_k(m)} \gamma_{k,m} \, q_k(m) \right] , \qquad (8)$$

where W_k denotes the optimum weight vector, λ_0 represents the noise floor, and $\gamma_{k,m} = q_k^H(m) \, S_T$ is the projection of the mth interference eigenvector onto the quiescient response. Observe that terms associated with the noise eigenvalues do not affect the optimum weight vector. This same observation is made in [6].

Other researchers provide alternative eigenbased interpretations. For example, the two methods proposed in [7] rely on constructing a weight vector lying in the noise subspace. Since the noise subspace is orthogonal to the correlated interference, selecting a weight vector $\hat{W}_k$ such that $\hat{W}_k \in span(Q_{noise,k})$ cancels correlated interference. The cross-spectral metric (CSM) method discussed in [8] applies a cost function to the problem of choosing the best low-rank eigenbasis. In this case, the weight vector generally spans the noise subspace as seen through the following alternative view of the CSM method. From (2)-(5), the signal-to-interference plus noise ratio (SINR) may be written for a normalized target signal as

$$SINR = S_T^H \hat{R}_k^{-1} S_T = \sum_{m=1}^{NM} \frac{| S_T^H \, q_k(m) |^2}{\lambda_k(m)} \; . \qquad (9)$$

Accordingly, with maximum SINR as the objective, a low-rank basis selection should choose those eigen-components which maximize the partial sum of terms in (9). The desired target response influences this selection. For example, consider the top plot in Figure 1 showing a typical eigenspectra for a simulated airborne radar clutter covariance matrix. In contrast, the bottom plot in the figure shows the individual "CSM" terms, $| S_T^H \, q_k(m) |^2 / \lambda_k(m)$, as defined in (9). It is evident from this example that terms with the largest CSM predominantly lie in the noise subspace.

4. Subspace Processing

The preceding section is central to the development of numerical approximations suited to the STAP problem. As just discussed, three choices emerge for constructing the reduced-rank weight vector: 1) $\hat{W}_k$ lies in the interference subspace; 2) $\hat{W}_k$ lies in the orthogonal noise subspace; or 3) $\hat{W}_k$ uses those basis vectors which optimize an objective function. If the basis vectors arise from the SVD, it is sensible to construct $\hat{W}_k$ from all principal components. Complications may arise if the interference rank is fuzzy.

An advantage of subspace processing is reduced sample support requirements for covariance matrix estimation. Generally, the airborne radar signal environment is inhomogeneous (eg., spatially varying clutter and discretes) and non-stationary (data is coherent over a limited time interval). These factors limit useful sample data. Subspace approaches appear robust in such cases and afford the possibility of localized adaptive processing schemes. To corroborate this notion, we offer Figure 2. This figure depicts the actual eigenvalues of a jammer covariance matrix (three jammers present) for a 16-element linear array. The top plot shows the principal components, whereas the bottom plot depicts the noise eigenvalues. Also included in both plots are the corresponding eigenvalues for the sample covariance matrix estimated from sample support using 1x, 2x and 10x the total degrees of freedom (DOF). The figure

shows the robustness of the estimation process in representing the interference subspace, whereas the full-rank uncorrelated noise is poorly estimated. The interference eigenvectors are also usually well represented with limited sample support.

Appealing aspects of subspace methods, specifically convergence to optimum performance and mitigated sample support, is offset by computational burden. Taking into account that actual measured data only approximately satisfies the statistical assumption of homogeneous observations, we find sufficient motivation for exploring approximate subspace methods.

Several other notions influence the development of such approximation procedures. First of all, we point out that $q_k \in span\{S_I(p)\}$, where $\{S_I(p)\}$ is the set of interference space-time steering vectors. For high interference-to-noise terms, the eigenbeams generally point in a single direction. However, the discrete Fourier transform (DFT) of the eigenvectors can produce results appearing like a difference pattern when interferers appear closely spaced, as is true for ground clutter returns. Thus, a linear combination of the $S_I(p)$ comprise each eigenvector and multiple terms can dominate. Secondly, asymptotic equivalence exists between Toeplitz and circulant matrices [9]. Ideal covariance matrices are Toeplitz, whereas sample covariance matrices are non-Toeplitz. Nevertheless, this implies the columns of a discrete cosine transform (DCT) or DFT may serve as appropriate surrogates for the actual eigenvectors of the covariance matrix. (DCT and DFT vectors are eigenvectors of circulant matrices.) Let G be a unitary matrix whose columns are selected DCT or DFT vectors, D be a diagonal matrix approximating the principal eigenvalues and E represent the residual off-diagonal terms. In the case of closely spaced interferers, we generally find that

$$\| \hat{R} - G\, D\, G^H \|, \qquad (10)$$

where $\| \bullet \|$ is an appropriate norm, is unacceptably large. A deterministic basis may not adequately diagonalize $\hat{R}_k$. An adaptive basis, generated via efficient numerical routines, offers the potential for better performance at modest cost.

5. Numerical Subspace Approximation

We may interpret STAP as the constrained minimization of output power subject to a linear constraint,

$$\min \hat{W}_k^H \hat{R}_k \hat{W}_k \quad such\ that \quad \hat{W}_k^H S_T = 1 . \qquad (11)$$

Define the data matrix of snapshots from (1) as

$$X^H = \left[X_1, X_2, \ldots, X_L \right] . \qquad (12)$$

Upon substituting the MLE for $\hat{R}_k$, (11) equates to the linear least squares problem (LLSP),

$$\min \| X \hat{W}_k \|_2 \quad such\ that \quad S_T^H \hat{W}_k = 1 . \qquad (13)$$

Assume that $\| S_T \|_2 = 1$ and let $H\, S_T = \pm e_n$, where e_n is the n^{th} column of the appropriately dimensioned identity matrix. Furthermore, let $H\, \hat{W}_k = u = \left(\mu_m \right)_{m=1}^{NM}$. We then express (13) as,

$$\min \| X H^H H \hat{W}_k \|_2 \quad such\ that \quad S_T^H H^H H \hat{W}_k = 1 \qquad (14)$$

from which we get the unconstrained form

$$\rho = \min_{\nu} \| Z_{n-1} \nu + z_n \|_2 ;$$
$$Z = X H^H = \left[Z_{n-1}\ z_n \right] ;\ \nu = \left(\mu_m \right)_{m=1}^{n-1} , \qquad (15)$$

where $n = NM$. For a linear solver based on a direct QR-decomposition, the cost of finding ν is $O(N^3 M^3)$ with L = 2xDOF. The required computational rate can be very large.

5.1 Low-Rank Approximation of the Data Matrix

The singular values of the data matrix relate closely to the eigenvalues of $\hat{R}_k$. When Z is of low numerical rank, as is often the case in airborne radar [3, 5-8], we can approximate (15) by another minimization problem in a lower dimension subspace $span(Y) \subset span(X)$ for which minimizer p' of

$$\min \| Yp \| \quad such\ that \quad S_T^H p = 1 \qquad (16)$$

ideally yields small residuals for $X\, p'$ except at target locations. If Y adequately represents the dominant subspace of X, and if the dimension of Y is considerably smaller than $dim(X)$, significant savings in computational cost result without sacrificing target detection performance.

The discussion of eigenbased methods suggests selecting Y via the dominant left singular vectors of the SVD, $X = U\, \Sigma\, V^H$ [5-6]. As already mentioned, the SVD is a very costly approach to rank-reduction. However, we may obtain a suitable low-rank approximation through bi-diagonalization of X. Furthermore, we do not have to complete the bi-diagonalization to extract meaningful information regarding the dominant subspace [10]. One accomplishes bi-diagonalization via sequences of left and right Householder transformations, U and V, applied directly to X such that $B = U^H X V$ is upper (or lower) bi-diagonal. After k steps of bi-diagonalization, it is with high probability that the singular values of $B_k = B(1{:}k,1{:}k)$ tend to be very good approximations to the largest singular values of X. Thus, when X is of low numerical rank we expect

$$Y_k = U_k(:,1{:}k)\ B_k\ V_k(:,1{:}k)^H \qquad (17)$$

to be close to the best rank-k approximation of X.

5.2 LLSP in Lower Dimensional Space

Consider the following subspace approximation, used in effect as a rank-reducing transformation to alleviate computational complexity. First, bi-diagonalize X until an appreciable gap is found between the 2-norms of column k of $B^{(k)}$ and column k+1 of $B^{(k+1)}$. Next, solve (16) in the subspace spanned by U_{k+1} alone,

$$\min \| B_{k+1}\ \tilde{p} \|_2 \quad such\ that \quad \tilde{S}_T^H \tilde{p} = 1\ , \qquad (18)$$

where $\tilde{S}_T = V_{k+1}\ S_T$ and $\tilde{p} = V_{k+1}^H\ p$. Note, the unitary operator U preserves measure and we select B_{k+1} after its removal. A solution for $\tilde{p}$ follows by placing the constrained minimization into an unconstrained form similar to (15). One can develop a recursive relation for $\tilde{p}$ to efficiently accomodate expanded approximate subspace dimension. Observe that $p = V\ [\tilde{p}\ ;\ \bar{0}\ ]$.

5.3 Refinement Via Conjugate Gradient

From (8) it is seen that the optimum weight vector satisfies $W_k \in span(S_T, Q_{PC,k})$. The formulation based on partial bi-diagonalization extracts information from the data matrix such that $\hat{W}_k \in span(S_T, \hat{Q}_D)$ where $\hat{Q}_D$ is an approximation to the dominant subspace. As expected, other sources of correlated interference not represented by $\hat{Q}_D$ influence target detection. When choosing a low dimension subspace, it is not necessarily true that the dominant terms most greatly influence performance. One must consider the cost function when ranking the importance of each subspace [8]. A pre-conditioned conjugate gradient (CG) iterative method applied to the partially bi-diagonalized system of equations allows us to refine the weight vector produced from the unconstrained minimization in the dominant subspace alone. This additional step effectively expands the approximate weight vector to include basis vectors representing the weaker sources of correlated interference. We use $[B_k\ ;\ \sigma_{min} I\ ]$ as an (n-1)x(n-1) preconditioner, where σ_{min} approximates the smallest singular values of B_k. The pre-conditioner reduces singular value spread of the partially bi-diagonal system. It is known that convergence is very good for the CG method when the data matrix is well conditioned [11]. Thus, CG iterations are applicable in this instance. The starting weight vector approximation is $[\tilde{p}\ ;\ \bar{0}\ ]$.

5.4 Computational Cost

The computational cost of the partial bi-diagonalization (step 1) is $O(L \cdot NM \cdot k)$, with k the number of bi-diagonalization steps. The pre-conditioned CG iterations (step 2) add $O(L \cdot NM \cdot G_C)$ computations, where G_C is the number of iterations.

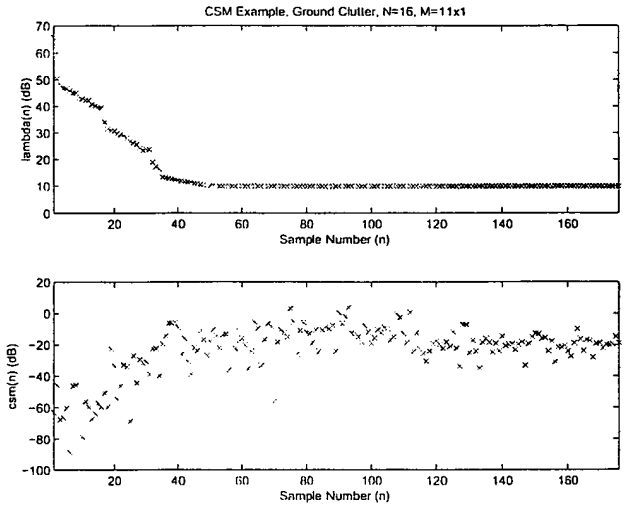


Figure 1. Eigenspectra and CSM terms.

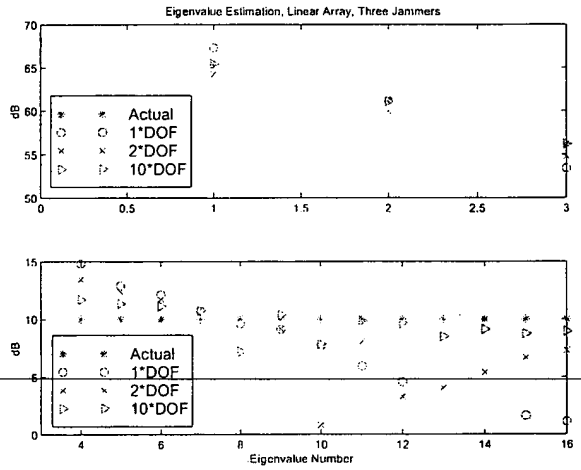


Figure 2. Robustness of dominant eigenvalues to low sample support.

6. Example

To validate the proposed method, consider the case of a 16 channel airborne linear array receiving 12 pulses. A weak target, positioned close to mainbeam clutter in Doppler space, is present in the 34th range bin (realization). Figure 3 shows the optimum filter output versus realization,

whereas Figure 4 shows the adaptive filter output using the conventional sample matrix inversion (SMI) [2]. In contrast, Figure 5 shows the result using the subspace approximation of partial bi-diagonalization followed by the pre-conditioned, unconstrained CG method. The order of the partial bi-diagonalization was k=16 and we used G_C = 18 CG iterations. Observe from the figures that all three methods detect the target. Interestingly enough, the approximate method gives the best results. We attribute this to improved numerical stability of the procedure in comparison to the numerical routines used to invert the covariance matrix for the optimum and SMI scenarios.

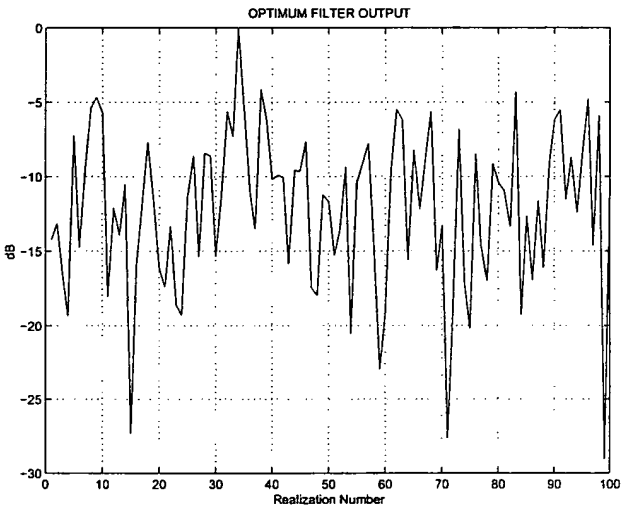


Figure 3. Optimum filter response.

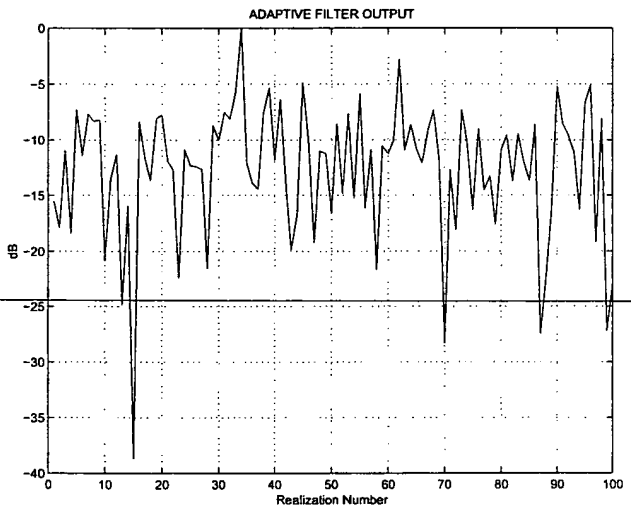


Figure 4. SMI adaptive filter response.

7. Summary

In this paper we propose an approximate subspace procedure suited to airborne radar STAP application. Recent eigenbased methods proposed by other researchers serve as motivation for our pursuit of this topic. The approximation involves partial bi-diagonalization and pre-conditioned conjugate gradient iterations to mitigate computational burden. A simple example shows that this approach has merit and warrants further consideration.

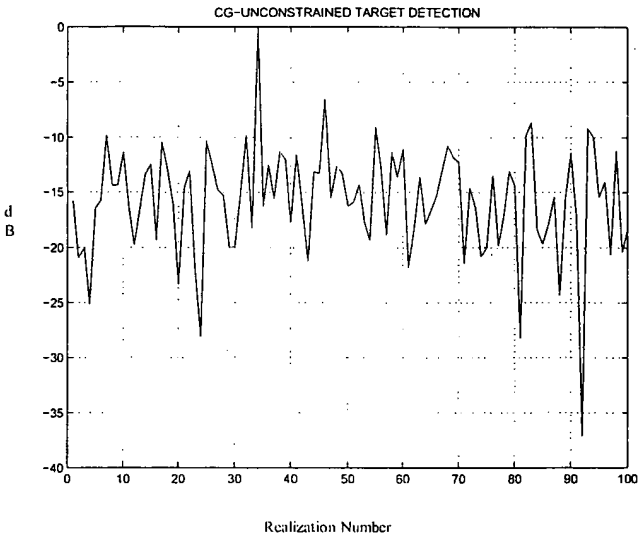


Figure 5. Unconstrained pre-conditioned CG method.

References

[1] L.E. Brennan and I.S. Reed, "Theory of Adaptive Radar", *IEEE Trans. AES*, Vol. 9, No. 2, pp. 237-252, March 1973.

[2] I.S. Reed, J.D. Mallett, and L.E. Brennan, "Rapid Convergence Rate in Adaptive Arrays", *IEEE Trans. AES*, Vol. 10, No. 6, pp. 853-863, November 1974.

[3] J. Ward, Space-Time Adaptive Processing for Airborne Radar, Lincoln Laboratory Tech. Rept., ESC-TR-94-109, December 1994.

[4] H. Wang and L. Cai, "On adaptive spatial-temporal processing for airborne surveillance radar systems," *IEEE Trans. AES*, Vol. 30, No. 3, pp. 660-670, July 1994.

[5] W.F. Gabriel, "Using Spectral Estimation Techniques in Adaptive Processing Antenna Systems," *IEEE Trans. AP*, Vol. 34, No. 3, pp. 291-300, March 1986.

[6] B.D. Van Veen, "Eigenstructure based partially adaptive array design," *IEEE Trans. AP*, Vol. 36, No. 3, pp. 357-362, March 1988.

[7] A. Haimovich, "The Eigencanceler: Adaptive Radar by Eigenanalysis Methods," *IEEE Trans. AES*, Vol. 32, No. 2, pp. 532-542, April 1996.

[8] J.S. Goldstein and I.S. Reed, "Theory of partially adaptive radar," IEEE Trans. AES, Vol. 33, No. 4, pp. 1309-1325, Oct. 1997.

[9] R. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," IEEE Trans. IT, Vol. IT-18, No. 6, pp. 725-730, Nov 1972.

[10] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd *Ed.*, Johns Hopkins Univ. Press, Baltimore, 1997.